

Tempo. 165590
NASA - 535

SECOND SUMMARY REPORT ON A REVIEW
OF BIOLOGICAL MECHANISMS FOR APPLICATION TO
INSTRUMENT DESIGN

DECEMBER 1963

DOCUMENT NO. ARA-T-9211-5 VOLUME I

PREPARED FOR :

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
BIOTECHNOLOGY AND HUMAN RESEARCH
OFFICE OF ADVANCED RESEARCH TECHNOLOGY
WASHINGTON 25, D.C.

OTS PRICE

XEROX

\$

MICROFILM

\$

19.75 ph

FACILITY FORM 602

N64-29546

(ACCESSION NUMBER)

315

(PAGES)

OR-58560

(NASA CR OR TMX OR AD NUMBER)

(THRU)

1

(CODE)

16

(CATEGORY)

SECOND SUMMARY REPORT ON A REVIEW
OF BIOLOGICAL MECHANISMS FOR APPLICATION TO
INSTRUMENT DESIGN



DECEMBER 1963

DOCUMENT NO. ARA-T-9211-5 VOLUME I

PREPARED FOR :

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
BIOTECHNOLOGY AND HUMAN RESEARCH
OFFICE OF ADVANCED RESEARCH TECHNOLOGY
WASHINGTON 25, D.C.

**SECOND SUMMARY REPORT ON A REVIEW
OF BIOLOGICAL MECHANISMS FOR APPLICATION TO
INSTRUMENT DESIGN**

December 1963

SECOND SUMMARY REPORT ON A REVIEW OF
BIOLOGICAL MECHANISMS FOR APPLICATION
TO INSTRUMENT DESIGN

by

J. Healer
H. Homonoff
C. Jamieson
D. Kimura
S. Koslov
J. Merchant
R. Papirno

Prepared for:

National Aeronautics and Space Administration
Biotechnology and Human Research
Office of Advanced Research Technology
Washington 25, D. C.

FOREWORD

The study presented in this report was performed for the Office of Biotechnology and Human Research, National Aeronautics and Space Administration under Contract No. NASw-535. The work was accomplished under the jurisdiction of Captain Frank B. Voris, MC USN, Chief, Biotechnology and Human Research Division and Mr. George Chatham, RBH, Office of Biotechnology and Human Research.

The program has been an interdisciplinary study of biological mechanisms, particularly biosensors, for application to instrumentation design and engineering and represents an extension of work done by Allied Research Associates, Inc. under Contract NASr-16.

TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	iii
LIST OF FIGURES	vii
SECTION 1 INTRODUCTION	1-1
SECTION 2 <u>MECHANORECEPTORS</u>	2-1
2.1 The Ear as a Transducer	2-1
2.1.1 Phylogenetic Development	2-1
2.1.2 Sound Energy Path	2-2
2.1.3 Output Signal	2-4
2.1.4 Displacement Sensitivity	2-7
2.1.5 Pressure Sensitivity	2-8
2.1.6 Mechanical Advantage	2-9
2.1.7 Ossicular Transmission Apparatus	2-9
2.2 Simulation of Aural Mechanical Safety Mechanisms in Pressure Transducers	2-12
2.3 Single Sideband Communication of Speech	2-12
2.4 Touch Receptors	2-14
2.5 Gravity Sensors	2-15
REFERENCES	2-17
SECTION 3 PHOTORECEPTORS	3-1
3.1 Introduction	3-1
3.1.1 Purpose of the Program	3-1
3.1.2 Device Applications	3-2
3.1.3 Information System Applications	3-4
3.1.4 Human Engineering Applications	3-4
3.1.5 Conclusion	3-6
3.2 Analysis of the Visual System	3-7
3.2.1 Introduction	3-7
Quality Sensors and Information Sensors	3-7
The Visual Mechanism of Information Sensing	3-9

3. 2. 2	The Quantitative Description of Luminous Energy	3-12
3. 2. 3	The Image Forming Process	3-26
3. 2. 4	Image Analysis	3-35
3. 2. 5	The Application of Quantum Considerations to Visual Sensing	3-40
3. 2. 6	The Sine Wave Representation of Optical Phenomena	3-44
3. 2. 7	A Theory of Peripheral Retinal Sensing	3-51
3. 2. 8	Signal Processing Techniques of Nature	3-54
3. 3	Engineering Applications	3-60
3. 3. 1	Introduction	3-60
3. 3. 2	The Applications of Oculometry	3-60
	A. Command Control	3-62
	B. Remote Eye Vision-An Enhancement of the Visual Sense	3-83
3. 3. 3	The Technique of Oculometry	3-102
	REFERENCES	3-112

SECTION 4	CHEMORECEPTION	4-1
4. 1	Mechanisms of Olfaction	4-1
4. 1. 1	Monolayer Theories	4-3
4. 1. 2	"Lock and Key" Concepts	4-6
4. 1. 3	Morphology of Respiratory and Olfactory Region	4-8
4. 1. 4	Electrophysiological Considerations	4-9
4. 1. 5	Radiation and Vibrational Theories	4-12
4. 1. 6	Membrane Adsorption and Penetration	4-17
4. 2	Insect Chemoreception	4-27
4. 3	Sensory Receptors in General	4-27
4. 3. 1	Mechanism of Drug Action	4-28
4. 3. 2	Electrochemistry of Membrane Potentials	4-29
4. 4	Applications	4-35
	REFERENCES	4-42

	<u>Page</u>
SECTION 5 THERMORECEPTION	5-1
5. 1 Warm and Cold Spots	5-1
5. 2 Warm and Cold Receptors	5-2
5. 3 The Thermal Stimulus	5-8
5. 4 Central and Other Influences on Sensor Performance	5-11
5. 5 Radiant vs Contract Stimulation	5-14
5. 6 Thermoregulation in Homothermic Animals	5-15
5. 7 Infrared Detectors in Snakes	5-19
5. 8 The Ampullae of Lorenzini	5-23
5. 9 Conclusion	5-25
REFERENCES	5-28
SECTION 6 THE NEURAL ASPECT OF SENSORY FUNCTION AND ITS APPLICATION	6-1
6. 1 Introduction	6-1
6. 2 Biological Discussion	6-3
6. 2. 1 The Nerve Axon	6-4
6. 2. 2 The Axon Membrane	6-39
6. 2. 3 The Logic of the Nervous System and Pattern Recognition	6-50
6. 3 Applications	6-80
6. 3. 1 Neuristor	6-80
6. 3. 2 The Biological Effects of Strong Magnetic Fields	6-89
6. 4 Summary	6-97
REFERENCES	6-99

LIST OF FIGURES

<u>Figure No.</u>		<u>Page</u>
2-1	Middle and Inner Ear	2-3
2-2	Mechanical Stimulation of Basilar Membrane and Resulting Electrical Potentials	2-5
2-3	Hair Cell Action during Deflection of Basilar Membrane	2-6
2-4	Action of Stapedius Muscle	2-11
3-1	Man Interacting with His Environment	3-5
3-2	C.I.E. Luminous Efficiency Curve	3-16
3-3	Pupil Diameter as a Function of Retinal Illumination	3-24
3-4	Approximate Relationship Between Trolands and Lumens/ Sq M	3-25
3-5	Transmission Factor of the Human Eye	3-27
3-6	Quantum Flux	3-31
3-7	Bandwidth of the Eye	3-34
3-8	Segment Coordinates	3-37
3-9	Segment Coordinates of Alphabetic Characters	3-38
3-10	Spectral Curves	3-41
3-11	Increment Threshold as a Function of Background Brightness	3-45
3-12	Formation of Diffraction Pattern	3-49
3-13	Typical Human Operator Arrangement	3-63
3-14	Command Control System with Oculometer	3-64
3-15	Typical Command Control System	3-65
3-16	Typical Fixation Area During a 30 Second Period of Time	3-67
3-17	Eye Motion During Steady Fixation	3-68

LIST OF FIGURES (Cont.)

3-18	Gain and Phase Relationships for Continuous Predictable and Unpredictable Target Motions	3-69
3-19	Target Acquisition	3-72
3-20	Pilot's Helmet	3-73
3-21	TV Guided Missile	3-74
3-22	Pilot's First View	3-77
3-23	Missile About to be Fired	3-78
3-24	Pilot's View After Missile is Fired	3-79
3-25	Pilot's View Just Before Missile Hits Target	3-80
3-26	Guidance Equation	3-81
3-27	Evasive Maneuver After Firing Missile	3-82
3-28	Distribution of Rods and Cones in the Human Retina	3-85
3-29	Simplified Schematic of the Retina-Brain Connection	3-86
3-30	Television System	3-88
3-31	Example of a Variable Definition Picture	3-89
3-32	Typical Inanimate Feedback Control System	3-91
3-33	Simple Method of Oculometry	3-104
3-34	Corneal Reflection	3-105
3-35	Measurement Method	3-106
3-36	Measurement Geometry	3-108
3-37	Pupil/Iris Boundary Definition - Photograph of Pupil Accommodated to Dim Light	3-110
3-38	Pupil/Iris Boundary Definition - Photograph of Pupil Accommodated to Moderate Light	3-110
3-39	Pupil/Iris Boundary Definition - Photograph of Pupil Accommodated to Bright Light	3-110
3-40	"Best Fit" Circle to a Set of Points	3-111
4-1	Section of Head to Show Nasal Cavity	4-2

LIST OF FIGURES (Cont.)

4-2	Electrophysiological Response to Various Odors	4-11
4-3	Penetration Path of Odorous Molecule Through Olfactory Cell Membrane	4-18
4-4	Accelerating Power of Various Odorants versus Olfactory Threshold	4-20
4-5	"Dislocating Power" of Various Odorants versus Their Cross Sectional Areas	4-25
4-6	Comparison of Observed Olfactory Thresholds with Calculated Thresholds	4-26
4-7	Vibrating Condenser Method for Measurement of Contact Potential Difference Between Two Parallel Plates	4-37
5-1	Examples of Krause's End Bulb	5-3
5-2	Nerve Ending of Ruffini	5-4
5-3	Section Through the Skin (corium)	5-7
5-4	Impulse Frequency of a Single Cold Fiber in the Cat Lingual Nerve	5-9
5-5	Impulse Frequency of the Steady Discharge of a Single Cold Fiber and of a Single Warm Fiber	5-10
5-6	Bolometer	5-12
5-7	Hypothetical Anatomical Connections of the Central Thermoregulatory System (A Closed Loop Servo System)	5-16
5-8	A Computer Model of the Thermoregulatory System	5-18
5-9A	Infrared Pit Sensor of the Viper	5-20
5-9B	Cross Sectional Diagram of a Pit	5-20
5-10	Schematic Diagram of an Infrared Moving Target Indicator	5-22
5-11	Cone of Reception of the Pit Sensor	5-24
5-12	Response of Ampulla of Lorenzini to Temperature Changes	5-26
6-1	Diagram of a Typical Adult Unmyelinated Nerve Fiber	6-5
6-2	High-power Micrograph of an Unmyelinated Nerve Fiber	6-6

LIST OF FIGURES (Cont.)

6-3	Diagram Summarizing the Formation of Myelin Sheath	6-7
6-4	Survey Micrograph of a Young Mouse Sciatic Nerve Fiber Showing Several Stages in the Formation of Myelin	6-9
6-5	Diagram of Medullated Nerve Fibers	6-10
6-6	The Various Structures Observed at Ranvier Nodes	6-11
6-7	Diagram of Ranvier Node Collars for Three Fiber Diameters	6-12
6-8	The Passage of a Nerve Impulse in a Non-Conducting Medium	6-14
6-9	Monophasic Spike Potential of Single Unmyelinated Axon in Oil	6-15
6-10	Derivation of Cable Equation for a Nerve	6-20
6-11	Electrical Circuit Representing Membrane	6-22
6-12	Rushton's Apparatus for Strength-Length Determination	6-28
6-13	Dependence of Excitation Threshold on Interelectrode Distance	6-28
6-14	Dependence of Excitation Threshold on Pulse Duration	6-30
6-15	Dependence of Chronaxie on Interelectrode Distance	6-30
6-16	Apparatus for Electrode Configuration Experiments	6-33
6-17	Excitation Electrode and Electric Field Configurations	6-35
6-18	The Acetylcholine Cycle	6-41
6-19	Finean's Conception of the Molecular Organization of the 170 Å Repeat Unit in Nerve Myelin (axon membrane)	6-44
6-20	Sodium Ion Current Through Axon Membrane as a Function of Depolarization Potential	6-46
6-21	Illustration of Oriental Thought	6-58
6-22	Illustration of Oriental Thought	6-58
6-23	Scan Method of Russian Character Identification	6-64
6-24	Scan Identification of Latin Characters	6-67

LIST OF FIGURES (Cont.)

6-25	Information Processed versus Information Gained in Scan Recognition System	6-70
6-26	Effect of Association Unit Removal on Trained "E"-"X" Discrimination	6-74
6-27	Cryogenic Superconducting Neuristor	6-82
6-28	Hall Effect Neuristor	6-83
6-29	Matched Line Attenuation in Cables	6-85
6-30	Membrane Sodium Ion Channel	6-95

1. Introduction

This report and the first Summary Report¹ summarize the results to date of a review of biological mechanisms for application to instrument design and engineering performed for the National Aeronautics and Space Administration, Biotechnology and Human Research, Office of Advanced Research and Technology, under Contract NASw-535. The report is published in two documents:

ARA-T-9211-5
Volume I

Second Summary Report on a Review
of Biological Mechanisms for
Application to Instrument Design

ARA-T-9211-5
Volume II

Bibliography of Biosensors, A
Sampling of the World Literature
1900-1963 (Third Edition)

Principles of biosensor organization, structure, and function, throughout the sensed parameters, were investigated for application to engineering problems.

Biological sensors are the mechanisms which enable a living organism to monitor its environment (both internal and external). As such, they are vital to successful survival. It is not surprising, therefore, that these tiny transducers and their associated systems are so sensitive and highly developed. In addition to their extreme sensitivity, they exhibit other characteristics desirable in instrumentation such as wide dynamic range, low power requirement, self-protecting mechanisms, specificity, etc.

Vital to the space program and technological progress elsewhere, the study of biosensors can also lead to a better understanding of man's sensory and data processing ability, thus making possible the development of new methods of communicating information into man, new transmission systems, new data processing techniques, solution to man/machine interface problems, and the incorporation of man, as a component, into a physical system.

The pressures and demands of our space exploration programs and the new requirements being established by our progress in space, are accelerating the need for advances in many areas of instrumentation. There is little assurance that all of these needs will be adequately met by the conventional process of refining or inventing strictly "physical" measuring devices and systems. It seems reasonable therefore, to examine the principles of a different technology--the technology of nature--which accomplishes many of our goals and to derive new techniques to solve the problems of today and anticipate the needs of tomorrow.

Some of the ways in which the functional mechanism of biosensors can be applied to develop new concepts of instrumentation, enhance and extend the human senses, and improve the sensitivity of existing instrumentation are illustrated in the remainder of this report and in the previous report.

The approach and method previously developed (see ARA-1025, Section I) as well as the body of data accumulated and evaluations made provided valuable guidelines and a base for this interdisciplinary work. The extensive literature gathering and review was updated and current data incorporated into the material analyzed. Potentially promising applications and directions recommended in the preceding report were investigated and developed further. Additional principles were investigated and applied.

Several of these concepts are ready for intensive investigation to demonstrate proof of principle and prototype development. Others need further study. And there are many yet to be uncovered by further engineering and biological analyses.

1	ARA-1025	Summary Report on a Review of Biological Mechanisms for Application to Instrument Design, January 1962
	ARA-1026	Bibliography on Biosensors, A Sampling of the World Literature, 1900-1961 (Second Edition)

2. Mechanoreceptors

From the point of view of engineering application, the structure and response of biological sensors detecting the mechanical characteristics of environment, orientation, pressure, etc., compared with other biosensors, can be most closely duplicated with available materials. There are, however, many design subtleties intrinsic in mechanoreception processes which can lead to entirely new, more sophisticated, or intrinsically simpler instrumentation to improve or replace currently available techniques for sensing mechanical parameters. For example, one very intensive area of physiological research is the mammalian ear. This is of interest to bionics because of its extreme sensitivity, methods of information handling and its self protective character. A more detailed and comprehensive discussion of physiological structure and function of other mechanoreceptors is presented in the first summary report on this program (see Ref. 2-1).

2.1 The Ear as a Transducer

Although the mammalian ear is a commonly accepted prototype of the microphone, it can also be regarded from an engineering point of view as a displacement detector and pressure transducer having characteristics of operation which are somewhat more sophisticated than those of its man made mechanical counterparts. It is with these features of its operation, rather than its function as a sound transducer, that we will here be concerned. Incorporated in its total structure are elements which serve as safety devices and filters as well as acoustic transmission elements. Man made pressure transducers, in which commonly the pressure induced deformations of an elastic member are sensed by electrical gages, do not incorporate specific mechanical safety devices or acoustic filters. In fact, safety margins for overpressures in man made instruments are by factors of two or three, while in the ear they may be as high as several orders of magnitude; any filtration which is required in man made instruments is performed on the electrical output signal.

2.1.1 Phylogenetic Development

Many of the invertebrates possess sensory mechanisms designed to respond to sound. Some of these have been described in a previous report

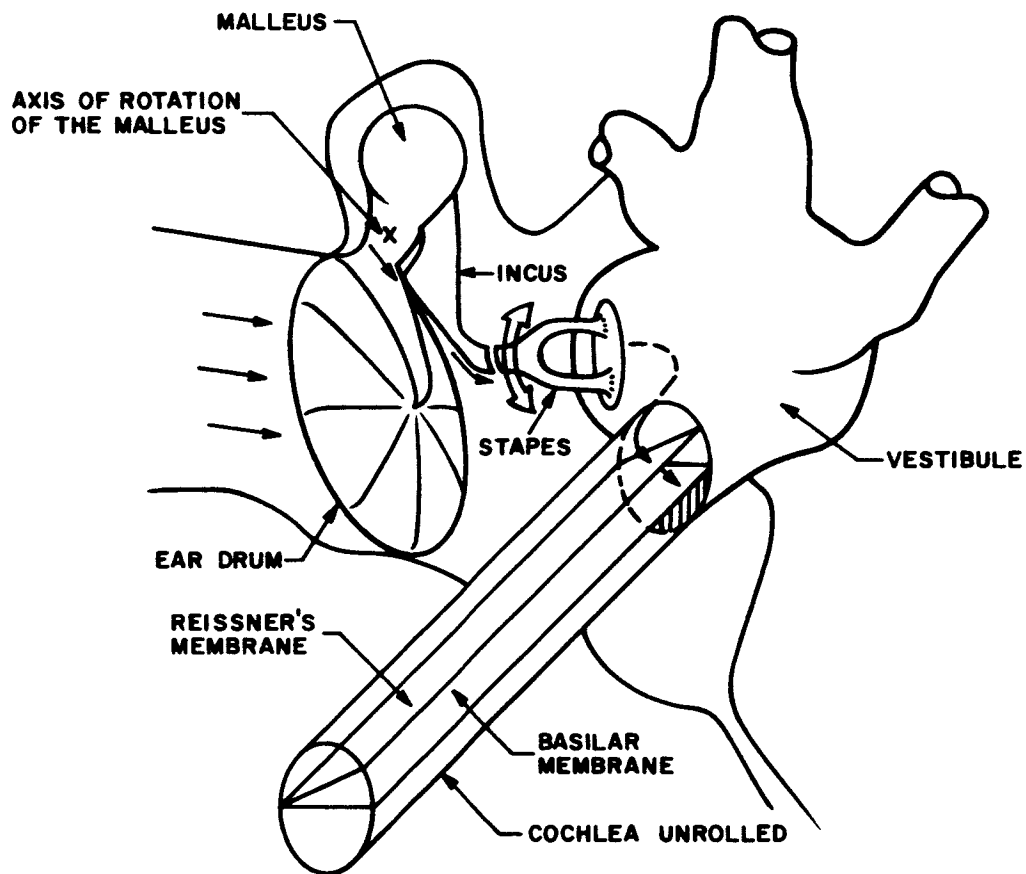
on this study ARA-1025, (Ref. 2-1). The first development of a hearing organ as such appears in the lowest vertebrates, the craniate fishes, in certain canals associated with the lateral line. It is in these animals that the hearing or vibration sensing apparatus is first removed from the outer environment by a canal structure with the hearing sensors being embedded in a fluid manufactured by the animal. Vibrations in the surrounding environment must first be transmitted through this fluid to the sensory cells. The canal structure of the lower vertebrates eventually evolved into the complex cochlea of the higher animals.

The auditory ossicles may have had their origins in the small ossicular structures associated with the swim bladders of the fishes. It is in birds that a single ossicle first appears to couple the external auditory structure with the cochlea sensing mechanism. In all the higher mammals the ossicular apparatus is virtually the same, consisting of the three ossicles forming the transmission line and accomplishing other subsidiary functions between the external tympanum and the internal cochlea.

2.1.2 Sound Energy Path

The path of acoustic energy impinging upon the human eardrum has been indicated schematically in Figure 2-1. The eardrum, a highly damped membrane structure, responds to the overpressures and underpressures by displacements which are transmitted to the malleus, the first of the three ossicles in the middle ear. These displacements result in rotations of the malleus which are transmitted to the second ossicle the incus. The latter, rotating in the same sense as the malleus, is in contact with the third ossicle, the stapes causing a rotation and a linear displacement of the latter. The foot plate of the stapes is in contact with vestibular fluid and inward stapes displacements cause compressions of the fluid to travel into the cochlea, a coiled structure which is shown in Figure 2-1 unrolled. The coiled cochlear passageway is divided into three compartments by Reissner's membrane and the basilar membrane. The fluid motions induced by the stapes occur in the upper compartments shown in Figure 2-1, are transmitted across Reissner's membrane, and result in a downward displacement of the basilar membrane. It is in the basilar membrane that the hearing organ is located and it has been

FIG. 2-1
MIDDLE AND INNER EAR SCHEMATIC,
ADAPTED FROM VON BEKESY, REF. 2-2



hypothesized that the displacement of the organ serves to trigger impulses to the brain via the auditory nerves to give the sensation of hearing. This explanation is admittedly much simplified but serves in a general way to indicate the path of sound energy transmission through the auditory apparatus.

2.1.3 Output Signal

It has not yet been possible to demonstrate the exact mechanism in the basilar membrane whereby the mechanical acoustic energy is transduced to an electrical signal. However, there are sensitive cells called hair cells that are surrounded by the endings of the auditory nerve. It has been hypothesized that pressure changes in the cochlea result in deformations of the basilar membrane which are transformed by the hair cells into electrical potential differences. Von Békésy has indicated that displacements of the hair cells result in electrical potentials called microphonics which are exactly in phase with the displacements (2-2). In experiments in which a trapezoidal mechanical displacement was applied to the basilar membrane, the potentials are also trapezoidal in form as shown in Figure 2-2. The results of these experiments indicate that the inner ear mechanism is a displacement transducer rather than a velocity transducer.

Stuhlmann (2-3) has suggested that the mechanical excitation of the hair cells occurs in the manner shown in Figure 2-3. The hair cells are embedded in the Organ of Corti and are in contact with the tectorial membrane. A displacement of the basilar membrane results in deflections of the tips of the hair cells and bending of the hair cell body. Stuhlmann suggests that the hair cells act as levers with the fulcrum at the surface of the cell where the bristle emerges. From an engineering approach, however, the cells can be considered to be individual cantilever beams and as such, the maximum stress in the deflected position occurs at the bristle junction point and this is transmitted to the hair cell base. Since the cochlear nerves are in contact with the hair cell bases, it would appear that the source of the microphonic potentials is in this area of maximum mechanical stress.

FIG. 2-2

**SCHEMATIC REPRESENTATION OF MECHANICAL
STIMULATION OF BASILAR MEMBRANE OF GUINEA PIG
AND RESULTING ELECTRICAL POTENTIALS**

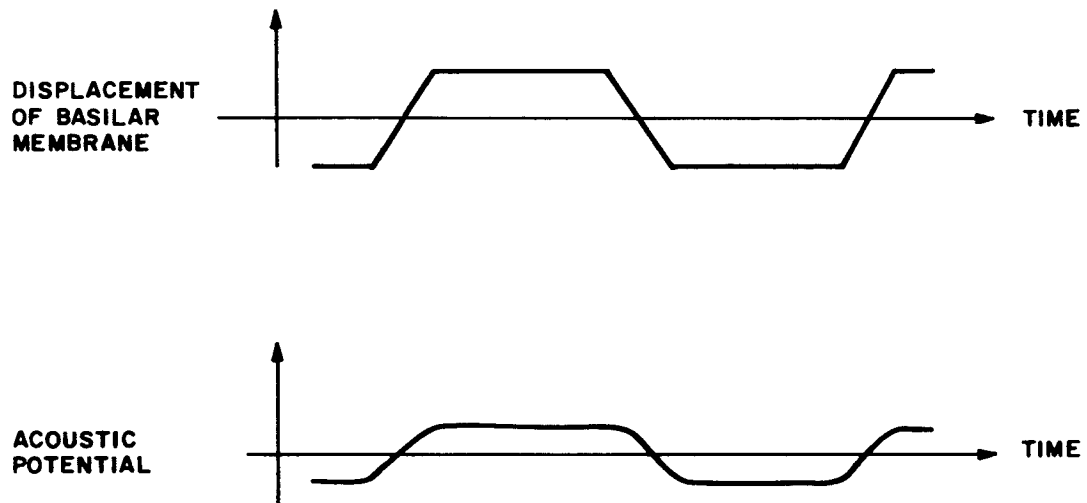
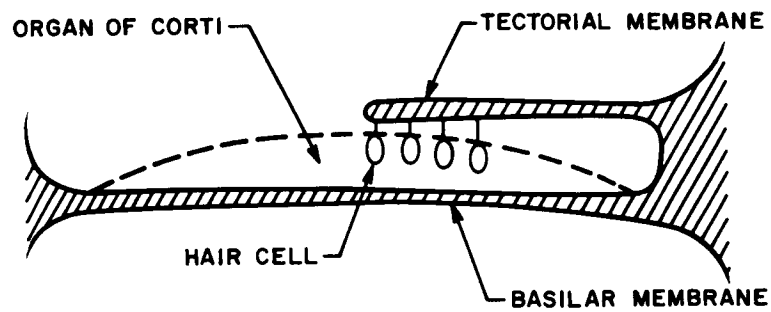


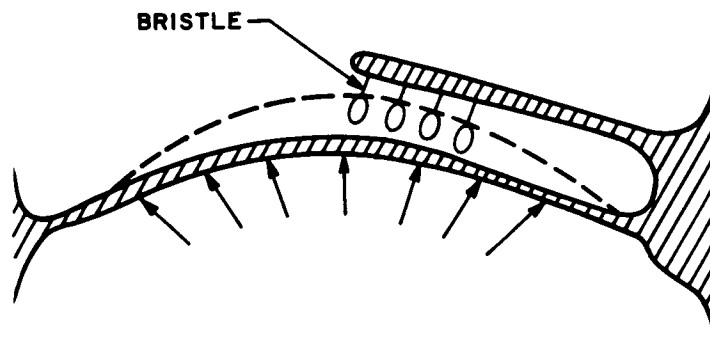
FIG. 2-3

SCHEMATIC REPRESENTATION OF HAIR CELL ACTION DURING DEFLECTION OF BASILAR MEMBRANE

a. UNDEFLECTED

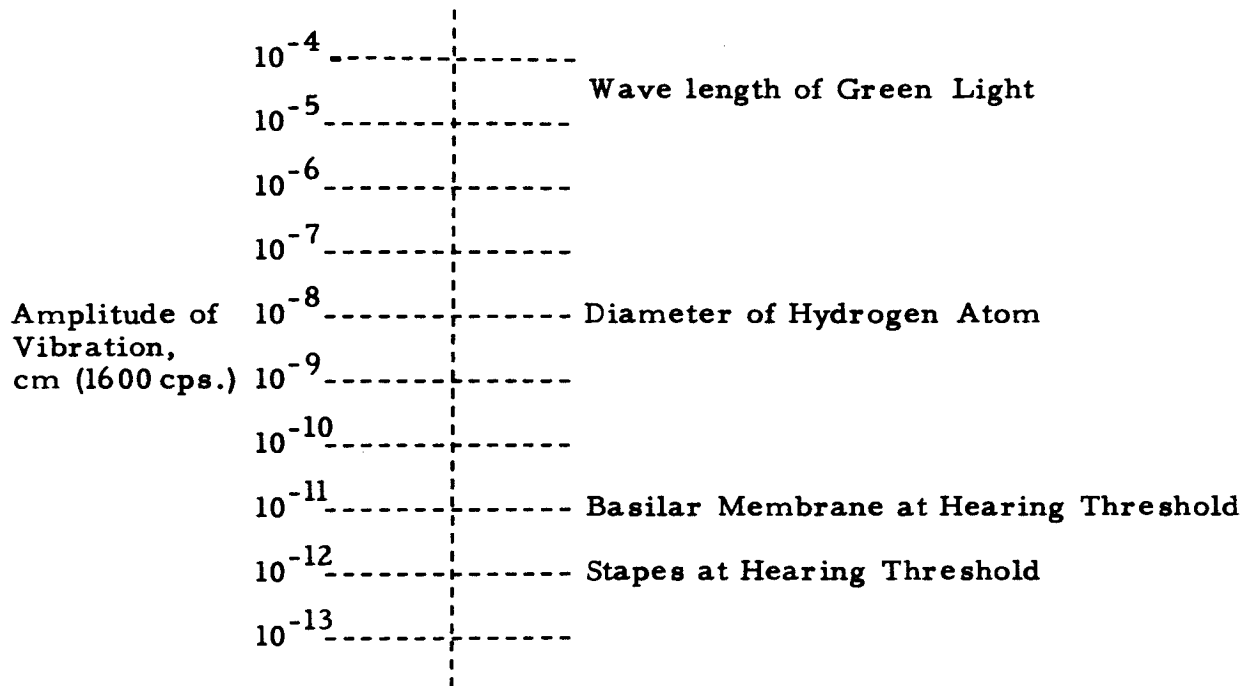


b. DEFLECTED



2.1.4 Displacement Sensitivity

The extreme displacement sensitivity of the auditory mechanism at the threshold of hearing make this structure an attractive one as a possible prototype of a dynamic transducer. It is of some interest to examine the quantitative values of the vibration amplitudes of inner ear structures at the threshold of hearing. At approximately 1600 cps, both the stapes and basilar membrane (see Figure 2-1) in man show the minimum displacement at threshold. These have been given below:



The threshold deflection values for the basilar membrane and the stapes appear to be slightly larger than the random amplitude which results from Brownian motion.

In considering the ear as a basic displacement transducer, it would be interesting to study mechanical methods of simulating the basilar membrane-hair cell complex as the sensing device for it is here that the extreme sensitivity of auditory apparatus probably arises.

2. 1. 5 Pressure Sensitivity

Examining the ear as a displacement sensing device, the set of structures associated with the basilar membrane was of interest. Examining the ear as a pressure transducer it appears to be more useful to consider a number of its elements rather than the cochlea sensing mechanism alone. These elements including the tympanum and the ossicles and their associated musculature and ligaments perform more than an energy transmission function, and it is with the associated ancillary functions which we will here be concerned.

In overall operation, the ear exhibits a large dynamic range of pressure sensitivity. At threshold the sensitivity of the mammalian ear to pressure variations is a function of frequency and is different for different species. Minimum values for a number of mammals is given in Table I below:

Table I. Auditory Threshold Pressures

	Minimum Threshold Pressure		Frequency to achieve minimum pressure cps
	psi	atmosphere	
Man	1.2×10^{-9}	8.1×10^{-11}	3500
Monkey	1.0×10^{-9}	6.8×10^{-11}	8000
Cat	7.2×10^{-10}	4.9×10^{-11}	8000
Guinea Pig	7.2×10^{-8}	4.9×10^{-9}	1000

The values of the minimum pressure have been given in atmospheres in Table I to emphasize the high sensitivity in terms of overpressure against base pressure since the ear operates at a base pressure of 1 atmosphere. In the human ear the threshold for pain is at an overpressure of approximately 10^6 times the threshold values, an exceptionally wide range.

2.1.6 Mechanical Advantage

Air pressure variations impinging upon the eardrum result in forces in the ossicles. These forces are applied over the area of the footplate of the stapes and cause pressure variations in the vestibular fluid. It is the pressure variations in the fluid which are sensed by the hearing organ, hence, a gross measure of the mechanical advantage of the system is the ratio of the vestibular hydraulic overpressure to that of a given aerodynamic overpressure at the surface of the eardrum. Experiments by von Békésy (2-4) have indicated that this pressure ratio may be as high as 14. The area of the eardrum in man is some 23 times the area of the footplate of the stapes, hence it appears that all the available area is not completely effective.

2.1.7 Ossicular Transmission Apparatus

In the mammalian ear, the three ossicles serve more than an energy transmission path. With their associated musculature and ligaments they perform protective and frequency filtering functions as well. The vestibular apparatus in the cochlea, as was shown previously, is enormously sensitive and very delicate and nature has provided several safety mechanisms to protect the sensing structure against dangerous overpressures. Small muscles attached between two of the ossicles and the walls of the middle ear cavity serve to change the acoustic coupling between the bones at their articulation and this mechanism allows filtering action of broad bands of high or low frequencies.

A major protective mechanism for high overpressures is the malleus-incus articulation. This joint can dislocate with large deflections of the eardrum, and re-articulate when the disturbance has subsided. There are two intra-aural muscles: 1) the tensor tympani which is inserted in the malleus, and 2) the stapedius, which is inserted in the stapes near its articulation with the incus. The action of the two muscles is opposed. Increased tension in both of the muscles tends to decrease the coupling between the malleus and incus, and thus allow dislocation to take place. In this mode of operation the malleus and incus act as a mechanical slip clutch which is put in the slip condition by muscle contraction. As such it may serve as a model for a mechanical safety device in a pressure transducer subject to

occasional excessive overloads.

It is interesting to note that the contractions of the tensor tympani and stapedius muscles bring additional safety mechanisms into play. The underpressure, following an overpressure of sufficient magnitude to cause malleus-incus dislocation, would tend to bulge the eardrum outward. The contracted tensor tympani restrains this action. This restraint is possible, since the malleus is firmly fixed to the eardrum in the region where the tendon of the tensor tympani is inserted into the neck of the malleus.

Contraction of the stapedius muscle increases the tensions in the membrane connecting the foot plate of the stapes to the periphery of the oval window. This action is illustrated in Figure 2-4, which has been adapted from Reference 2-2. Von Békésy (2-2) has estimated that the contraction of the stapedius in man can attenuate lower frequencies by a factor of 2 which in effect halves the mechanical advantage of the composite system.

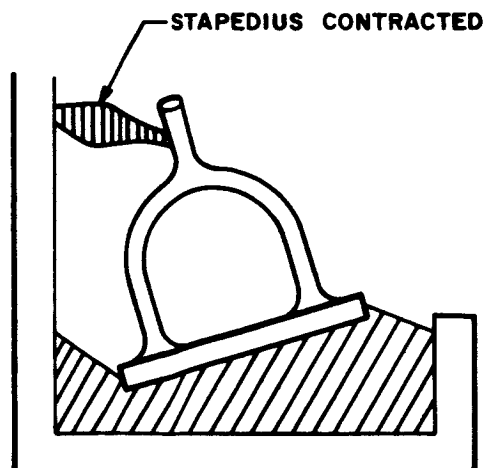
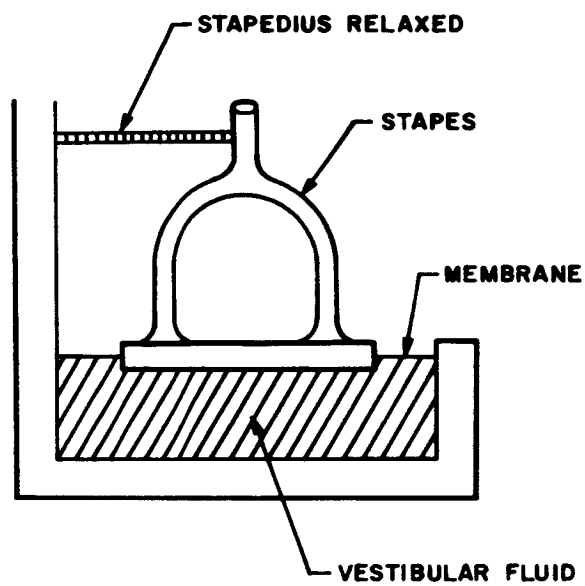
The stimulus for the intra-aural muscle contractions has been shown to arise in the eardrum itself (2-5). The muscle contractions are a reflex response to excitation of the eardrum.

Varying tensions in the two muscles also serve to alter the transmission characteristics of the system. As tension in the muscles increases, the lower frequencies become more attenuated, and transmission of the higher frequencies is enhanced. This action may be characterized as a push-pull effect with suppression of the low end of the sound spectrum coupled with the simultaneous amplification of the high end. At least part of this effect is due to the change in coupling at the malleus-incus articulation.

The ear may serve a model of a high sensitivity pressure transducer which employs redundant mechanical safety devices for high overpressure and underpressure conditions. It incorporates mechanical frequency filtration and both the safety mechanisms and filters are actuated in external elements. This in effect protects the sensitive sensory mechanism from damage prior to the transmission of dangerously high pressures to the cochlea.

FIG. 2-4

**ACTION OF THE STAPEDIUS MUSCLE IN
INCREASING THE TENSION OF THE STAPES
MEMBRANE WHICH DECREASES THE
TRANSMISSABILITY OF LARGE PRESSURE PULSES
INTO THE VESTIBULAR FLUID. (SCHEMATIC)**



2.2 Simulation of Aural Mechanical Safety Mechanisms in Pressure Transducers

The main safety mechanism against overpressure in the mammalian ear is in the ossicular chain - the three middle ear bones and their associated muscles, ligaments and articulations. The application of this principle to the development of a practical, self-protective pressure transducer must first delineate the mechanical principles involved in protecting the mammalian hearing sense organ against overpressure then design and construct a mechanical model of a pressure transducer, which incorporates the safety principles found in the biological counterpart and the scaling factors determined necessary for the application. The investigation should consider the effects of the incorporated safety mechanisms on the transmitted dynamic pressure signal. It would be desirable to determine signal attenuation as a function of frequency. Although considerable material exists in the literature the operation of the biological mechanism in its safety mode must be detailed by dissection of middle ear preparations of large and small mammals as well as enlarged mechanical models of the mammalian ear structure.

We may then proceed to the design of a breadboard model of a pressure transducer to automatically withstand dynamic overpressure of at least one order of magnitude over the maximum measurable pressure. The actual pressure range of this transducer would be determined and the operation of the designed safety mechanism tested.

2.3 Single Sideband Communication of Speech

In radio communication, speech is transmitted by the modulation of a carrier wave. In amplitude modulation, the amplitude of the carrier wave is modulated by the speech waveform. Thus let f_c be the carrier frequency, 100 a the percent modulation, and $f(t)$ the speech waveform. The modulated carrier wave may be expressed as

$$F(t) = [\sin 2 \pi f_c t] (1 + a f(t))$$

Suppose that $f(t)$ is a pure sine wave of frequency f_m . Then,

$$\begin{aligned} F(t) &= [\sin 2\pi f_c t] (1 + a \sin 2\pi f_m t) \\ &= \sin 2\pi f_c t + \frac{a}{2} \left\{ \cos 2\pi (f_c - f_m) t \right. \\ &\quad \left. - \cos 2\pi (f_c + f_m) t \right\} \end{aligned}$$

Thus, as a result of the amplitude modulation, the radiated wave $F(t)$ now consists of the original unmodulated carrier wave ($\sin 2\pi f_c t$) together with two "sidebands" - i.e. waves whose frequency is displaced by an equal amount of f_m above and below the carried frequency f_c .

In single sideband transmission only one of these three components of the modulated carrier is transmitted. This may be either the upper ($\cos 2\pi (f_c + f_m) t$) or lower ($\cos 2\pi (f_c - f_m) t$) sideband. One reason for doing this is to conserve transmitted power. Theoretically all the intelligence information is contained in one sideband and thus there is no need to transmit the other sideband and the carrier.

In order to rederive the speech intelligence at the receiver, the carrier must be reinjected. This may be done by product detection: the received signal, say $\cos 2\pi (f_c + f_m) t$, is multiplied by the reinjected carrier $\cos 2\pi f'_c t$, yielding

$$\begin{aligned} \phi(t) &= \cos 2\pi (f_c + f_m) t \times \cos 2\pi f'_c t \\ &= \frac{1}{2} \left\{ \cos 2\pi (f_c + f'_c + f_m) t \right. \\ &\quad \left. + \cos 2\pi (f_c - f'_c + f_m) t \right\} \end{aligned}$$

f_c and f'_c will be almost equal. Let $\delta = f_c - f'_c$, i.e. the error in the reinjected carrier frequency.

The result of the product operation may be filtered to remove the high frequency component ($f_c + f'_c + f_m$), yielding a detected signal

$$\psi(t) = \cos 2\pi (f_m + \delta) t$$

The original intelligence at the transmitter was at frequency f_m , so the result of the whole operation is a frequency shift δ . When the intelligence is a speech band, ($f(t)$), rather than a single frequency, the result is a shift of all the frequency components of $f(t)$ by a fixed amount δ .

In practice the effect of this carrier error is quite remarkable. For values of δ greater than about 100 cps the speech becomes unintelligible. For δ equal to 5 cps a very noticeable change in speech quality occurs.

Automatic Frequency Control Requirements

The practical significance of this effect is that the reinjected carrier frequency must be made equal to the transmitted carrier frequency to a very high degree of accuracy. Very precise crystal controlled oscillators can--in general--be used to obtain the necessary frequent control. However, in the case of communication between a moving aircraft and the ground, the Doppler Shift can be significant. For example, consider an aircraft flying at Mach 1. The Doppler Shift will then be approximately 1 part in 10^6 . For a carrier frequency of 100 mc/s this would amount to a 100 cps frequency shift. Thus because of the Doppler Shift, single sideband communication between an aircraft flying at this speed relative to a fixed station at 100mc/s will be difficult. The reinjected carrier would have to be manually adjusted wherever the relative velocity changed. A human operator can easily adjust the carrier frequency, by tuning for maximum intelligibility or clarity. It would be very desirable to have an automatic device that could perform the same function.

This suggests a very useful area of bionic research. It is an experimental fact that man can detect very small carrier errors. If the aural mechanism is analyzed to find out how this is accomplished, the basis may be discovered for a practical electronic frequency corrector.

2. 4 Touch Receptors

A particularly interesting feature of some touch receptors such as those constituting the cat's paw is the phasic mode of information transmission. In its simplest definition the phasic mode is a single pulse or pulse train of intensity independent of the magnitude of the event being transmitted; opposed to a tonic transmission mode where a fixed frequency is transmitted for some time period.

In the case of the paw receptors, specific areas are almost independently sensitive to the contact. However, there are some overlapping areas of sensitivity. Thus a weak pressure may only activate, say, a single receptor, while a strong pressure may be transmitted to nearby receptors as

well. The information handling system of the organism can then determine relative strength by the number of receptors affected. A detailed analysis of this mode of behavior involves the transmission of a compressional wave along the surface of the sensing pad. The information handling system can then define specific characteristics of the type of touch involved, i. e. something sharp or dull, by the time of propagation and attenuation of the propagated pulse to distant receptors of the surface. From the point of view of application to an instrument concept, this all or nothing phasic response would appear to lend itself to the development of a device for determining surface characteristics. An example might be a moon surface probe to determine what the characteristics of the surface materials are, hard or soft, small or large, and possibly also penetration depth for a given impact velocity.

Another possible contact sensing device is being considered based on the bioelectrogenic process discussed in Sections 4 and 6, i. e. using a variation in permeability of a membrane to specific ions, where the variation of permeability is caused by mechanical stretching of the membrane. In the normal neural process, the change in permeability is produced by chemical action (release of acetylcholine). The electrical signal is produced by the ion currents flowing in response to the permeability change. The electrical pulse is time limited again by a chemical process, the enzymatic hydrolysis of the acetylcholine restoring the ambient permeability, and the original ion balance is restored.

2.5 Gravity Sensors

Examination of the orientation sensing mechanisms of both invertebrates and vertebrates can lead to interesting concepts for simple, high reliability although probably low precision devices.

One concept which could be easily engineered in detail developed from examination of the statocyst of the crustaceans discussed in ARA 1025 (Ref. 2-1). It is an angle sensing device for a zero gravity environment, which would permit a simple and inexpensive method of determining the orientation angle of some device relative to some arbitrary given direction. The arbitrary direction is established by a small strong magnet. The sensing device, is a small ferromagnetic and conducting sphere, free to roll.

on the inside of a larger electrically insulated sphere. This larger sphere would have a grid of double wires crossing in longitude and latitude, with the spacing between the wires such that the conducting ball would create a short. These wires would, of course, have suitable insulation division points. A simple low-voltage battery then could be the sole power source, and each double wire section can be connected to a suitable display tube unit to indicate what position the ball has come to rest in as a result of the magnetic field attraction thus indicating the angle. These wires are, of course, analogous to the cilia of the statocyst, and the steel ball to the small object contained in the crustacean statocyst.

The semicircular canals of the mammalian ear as an orientation sensor can be directly taken over into a magnetohydrodynamic device. Consider a circular tube of a non-conducting material with a low surface friction, say Teflon. If a magnetic field is applied perpendicular to the conducting fluid, then any rotation of the tube will induce a current in the mercury. This current in turn can be detected by a separate sensor coil and used as a measure of the degree of rotation.

REFERENCES - SECTION 2

- 2-1. Healer, J., Arms, A., Merchant, J., and Britt, E., "Summary Report on a Review of Biological Mechanisms for Application to Instrument Design," ARA-1025, Jan. 1962.
- 2-2. von Békésy, George, "The Gap Between the Hearing of External and Internal Sounds," in Symposia of the Society for Experimental Biology, Number XVI, Biological Receptor Mechanisms, Academic Press, N. Y., (1962), p. 267-288.
- 2-3. Stuhlmann, Jr., Otto, "Hearing-Mechanism," Medical Physics, Vol. II, Otto Glasser, Ed., Yearbook Publishers, N. Y., (1950), p. 385-399.
- 2-4. von Békésy, G., "Über die Messung der Schwingungs-amplitude der Gehörknöchelchen mittels einer kapazitiven Sonde," Akust. Zeit., 6, 13, (1941)
- 2-5. Wiggers, H. C., "Functions of the Intra-aural Muscles," Am. J. Physiol., 120, 771, (1937)

3. Photoreceptors

3.1 Introduction

3.1.1 Purpose of the Program

Practical engineering applications can be derived from a study of living sensory mechanisms -- in this case, visual sensory mechanisms.

There are three possible classes of engineering applications in this study of vision. These may be termed "device" applications, closed loop information system applications, and human engineering applications.

Device applications involve the utilization of particular features of the living world for physical instrumentation. They result from the translation of principles of biodesign into engineering design. Were it not already invented, the camera could be cited as a composite example of several such applications, i. e., the lens, the iris aperture, the closed box frame of the camera (cf. eyeball) and the screen at the back of the camera (cf. retina).

Information system applications involve physical modeling of the total control system whereby an organism visually senses its environment, processes this data, and then generates useful motor actions in, and on, its environment. For example, a bird of prey constantly scans its environment for a suitable "target", and when this has been located it is attacked. An example of an information system application would be the modeling of this process in a missile system. The missile would be fired before a specific target had been located by the human operator and it would then search out a suitable target automatically. The attack mode of operation would be initiated, again automatically, only when such a target had been located. The Sidewinder is an example of a very simple missile system of this sort. System analogs of it in the living world can be seen in the very primitive life forms that can merely sense the presence or absence of radiation. However, a bird, in picking out and attacking its prey, performs a far more sophisticated sensing operation. A very desirable engineering development would be the modeling of this vastly superior sensing capability in a missile system -- which could then, for example, be directed at the general direction of enemy and would home-in on suitable targets -- such as tanks -- whenever it "saw" them.

Human engineering applications involve an improved coupling of the human visual system to the "machine" at the man/machine interface. These

applications will result from an engineering systems analysis of the visual system. Some of the various applications of oculometry (the measurement of eye motion) are pertinent examples of this class of application.

3.1.2 Device Applications

Physical counterparts exist for parts of the visual system. For example, lens, pupil, refractive surfaces, etc., -- the problem is that much of the visual process is not well enough understood to generate instrumentation based on it.

The processes leading up to and including the formation of the image on the retina seem to be well accounted for in all important respects. Beyond this, very little appears to be known. For example:

1. How is the photon energy transformed into the energy of neural activity? This can be considered in two parts.
 - a. Initial transformation of photons into some form of chemical energy.
 - b. The transformation of this energy into neural activity.
2. How is color information extracted from the image? It may be added that the exact nature of the color information actually sensed is not completely understood. We may refer, for example, to the dichromatism of the fovea, the two color demonstrations of Land, and to recent experiments with stabilized color images on the retina. The great mass of experimental data on the characteristics of normal and abnormal color vision would certainly seem to indicate some form of tri-stimulus mechanism. An intensive search has been made for three photopigments with different absorption characteristics. This has not been completely successful and it has been suggested that color resolution may occur within each individual cone. The dimensions of the cone are such that it could function as an optical wave guide, and the standing wave spatial pattern could be the source of the color information.

3. It is a matter of great surprise that the physiology and psychophysics of vision is devoted almost entirely to foveal vision and hardly at all to peripheral vision. Yet, as an information channel, peripheral vision may be far more important than foveal vision. Not only is there very little quantitative data on the characteristics of peripheral vision, there is apparently, a lack of any clear description, in qualitative terms, of what the peripheral retina sees. The peripheral retina almost certainly performs a useful data processing function. How this is done and what exactly is the mathematical nature of the processing is not known.
4. The detailed mechanism by which the conscious image is stabilized, independent of eye, head or body motion is not known. When for example, the eyes are moved in their sockets, the retinal image moves relative to the photoreceptors -- yet there is no sensation of motion of the image. If however, the eye is kept still and the external scene made to move in such a way as to cause an identical relative motion between the retinal image and the photoreceptors, an acute sensation of motion of the image will result. In both cases the retinal stimulation is identical, yet the conscious effect is quite different. The implication, of course, is that there is proprioceptive sensing of eye motion (as well as head and body motion) and this information is used by the brain to interpret, or in TV parlance, "to sync" the data coming from the retina.

Not only is there a dearth of detailed information concerning the operation of many parts of the visual system, but also most of the potentially interesting features of the eye depend on the unique molecular system structure of living things different from our currently macroscopic technology. In general terms the following features of the eye would appear potentially interesting for application, as advances in molecular technology may permit.

1. Simultaneous execution of a large number of logical operations in the retinal networks that connect the photoreceptors to the optic nerve.

2. Molecular circuitry that will allow for the reception, and subsequent transmission over several inches, of signals made up of between 1 and 10 photons.
3. Logical operations in the retina responsible for the special characteristics of peripheral vision and also of dark adaptation.
4. Optical wave guide phenomena, leading possibly to color sensitivity.
5. The system circuitry that permits stabilization of the conscious image, independent of head or body motion.

3.1.3 Information System Applications

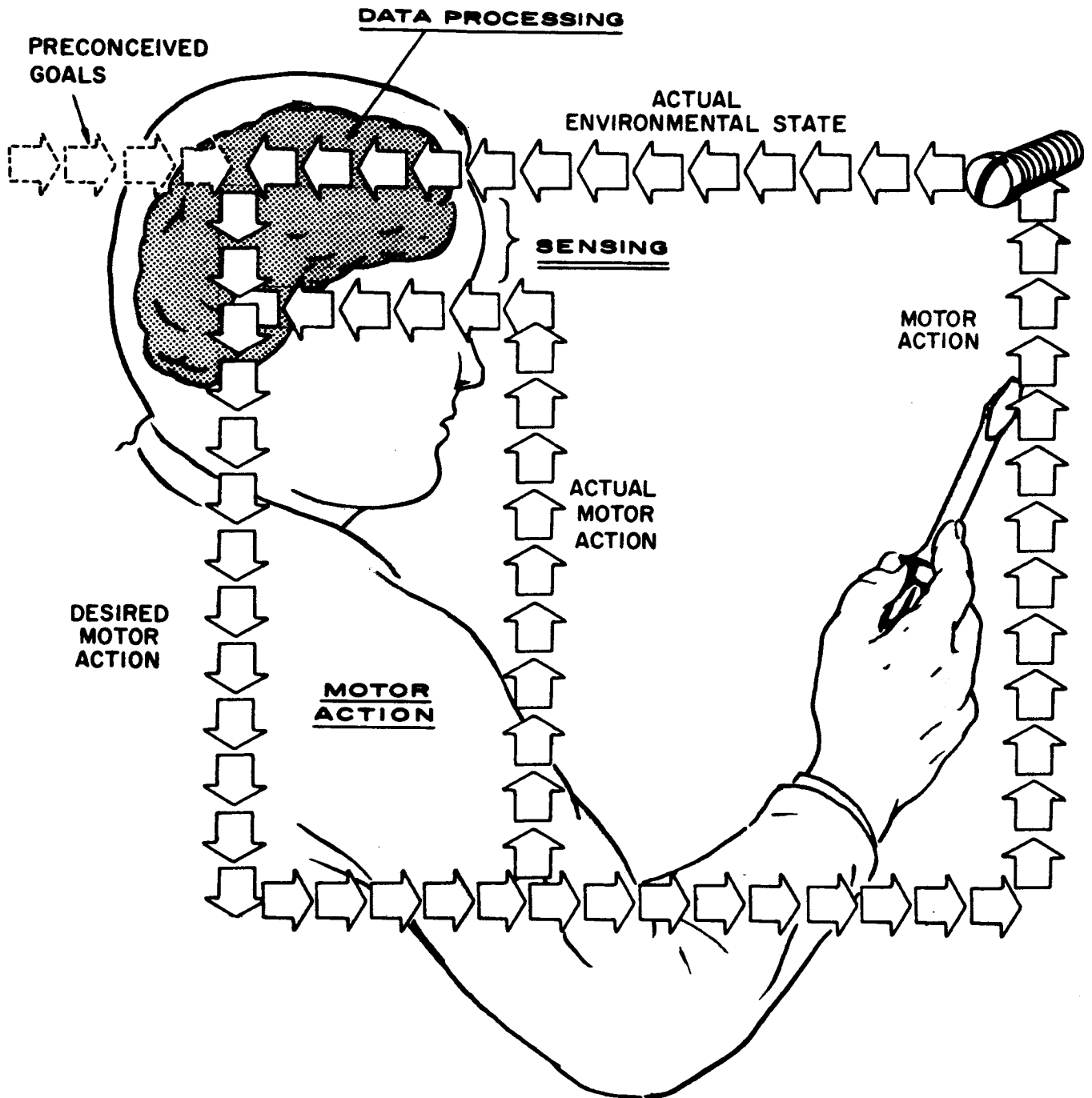
Let us examine an organism, for example a man, in normal contact with his environment. He receives information from his senses and generates motor action with his muscular system. The system is illustrated in Figure 3-1. The life blood of this system diagram is information. The senses (in particular the eye) absorb information and the muscles "return" information to the environment. We are concerned here with duplicating the ability of a man (or other organism) to function in this closed loop contact with the environment. We would like first to be able to describe the nature and quantity of visual information sensed. This is generally, naively, assumed to be some sort of point-by-point brightness description of the visual scene -- e.g. just like T V. As discussed in Section 3.2.7 and 3.2.8 this is almost certainly not true for anything except the narrow beam of foveal vision (in man). The fact that visual signal processing -- perhaps according to the spectral sensing model discussed in Section 3.2.7 -- can supply an organism with enough information to operate in its environment, suggests strongly that we should investigate these techniques for application to optical pattern and form sensors, etc.

3.1.4 Human Engineering Applications

Human engineering is concerned with the efficient integration of man into machine systems. Man can interact with these systems only through his senses and his motor functions. It is natural therefore that a study of a human sense -- in this case vision -- should lead to human engineering applications.

FIG. 3-1

MAN INTERACTING WITH HIS ENVIRONMENT



Recognition of the great sensory capacity of the peripheral retina and also of the central -- not incidental -- importance of the oculomotor system of the eye, has led to the conception of a range of human engineering system applications. Each involves a machine counterpart of this very important part of the visual system. This is an oculometer -- an instrument to continuously measure the angular orientation of the axis of regard. The use of this instrument -- as a link between man and machine -- can result in a substantial improvement in the match between his unique visual system and the "machine" with which he interacts.

3.1.5 Conclusion

Previous work on vision under this program (see Ref. 3-1) concentrated on oculometry and its potential human engineering applications. These have now been developed somewhat further and progress has also been made in the design of the necessary instrumentation (oculometer). This particular class of practical applications has appeared, from the very first, as an extremely fruitful and important result of this bionics study.

New work has been devoted essentially to information system applications. Ultimately it is desired to incorporate in a physical system the ability of an organism, using its visual sense, to function in an environment. It can be said only that a beginning has been made in this work. Attention has been devoted to the amount and kind of information sensed by the eye.

Device applications, as we have defined them, of photoreceptors, need further investigation as indicated in Section 3.1.2.

3.2 Analysis of the Visual System

3.2.1 Introduction

This is a study of photoreceptors regarded as information sensors or information transducers.

New principles, or system techniques, may be discovered that can then be applied in our man made systems that must be designed to bridge physical limitations and operational requirements.

The study involves the application to physiological detail of the principles of information theory and the techniques of system analysis directed to the goal of gaining engineering ideas and improvements.

The approach followed involves analysis of the visual information available to the sensor and the nature and quantity of the information absorbed by the sensor.

Quality Sensors and Information Sensors

A biosensor (like any sensor) not only senses a physical quality - e.g. heat, light, sound, etc. - but also absorbs information from the sensed environment. The information can be thought of as a modulation imposed on the physical quality. Thus, a radio receiver senses electromagnetic waves - the information carried by the waves is extracted and may then be transmitted in a context quite distinct from the original electromagnetic wave - e.g., as sound waves in air. To emphasize the distinction further between "information" and the physical quality with which it may (temporarily) be associated, we may note that a photomultiplier tube and an image orthicon TV camera tube are both light sensors, but the photo tube can detect only one unit of information, whereas the camera tube may detect millions of similar information units.

There are, therefore, two quite distinct ways of studying biosensors with a view to gaining ideas and new principles for engineering developments. We may either study the biosensor as a sensor of the physical quality, or we may study it as an information transducer.

The former approach often reduces to chemistry. Thus the initial detection of light energy in bio-photo-receptors is photochemical. This general result is not surprising, since the primary method that

living systems use to transmit information from point to point is chemical (i. e., nerve fiber propagation), thus there must at some point or points, in the chain, be a transformation from the physical quality (e. g., light) to a chemical effect.

A study of chemical detection methods could result in an extension of the range of materials available for component design. For example, photo-tubes now employ a range of photoelectric metals such as caesium. It may be possible to duplicate the very great dynamic range of bio-photodetectors by using, instead, photochemicals. Sensitivity in itself may not be of the greatest interest, since many transducers now existing can be made to operate at, or near, the ultimate physical limit of sensitivity. For example, the quantum efficiency of photocathodes can be as high as 10% - which is about as good as the human retina. It is impossible, from an engineering point of view, to divorce consideration of the sensor from the means used to carry information from it to the rest of the system. Thus development of chemical detectors - as analogues of biodetectors - may require a parallel development of chemical transmission circuits analogous to nerves. It may well be then that the main benefit of chemical sensors (of light, pressure, etc.) will derive from the possibility of subsequent chemical transmission of information from the sensor. Chemical transmission of information, or to generalize, transmission by means of neuristors, (which may or may not be chemical in nature) could have the following advantages over electronic conduction:

1. Very thin "wires" may be used. An electric signal sent along a very thin copper wire soon gets lost in the thermal noise.
2. Very complex, large scale, simultaneous logic operation become possible.

As an example of the importance of these factors, consider how an electronic model could be made of the human visual system. The retina absorbs, and the optic nerve transmits, information at the rate of about $10^7 - 10^8$ bits/second. It is probable that the data processing associated with visual perception involves very large scale processing of the data detected by the photoreceptors. This is accomplished in the

power) and liminal brightness increments define the information sensing capability of foveal vision. The validity of this descriptive of foveal vision can be considered as well established. It is demonstrated by the fact that when a picture is presented to the fovea with an information content equal to the maximum foveal information sensing capabilities, further improvement in picture quality is virtually undetectable. However, for peripheral vision these simple descriptives--acuity and liminal brightness increments--are quite inadequate to describe the visual process. This can be demonstrated by measuring foveal and peripheral (grating) acuity and presenting test letters foveally and peripherally that subtend 5 times the grating acuity in each case. The test letters presented foveally in this way are easily identifiable. If peripheral vision was of the same nature as foveal vision--i. e., described quantitatively by the same parameters--then the peripheral presentation of test letters should be equally clear, since the ratio of the size of the letters to acuity has been kept constant. However, it will be found that the letters are totally unidentifiable in the peripheral presentation.

The conclusion is reached that there is a lack of knowledge of the amount of information observed peripherally because the very nature of the peripheral sensation is not known. Mathematically, peripheral vision must respond, not--like foveal vision--to the brightness pattern itself (up to some spatial cut-off frequency), but rather to some function of it. When the nature of this function is known, it may then be possible to estimate quantitatively the amount of information absorbed by peripheral vision.

A hypothesis is presented in Section 3.2.7 concerning the nature of this function of the scene brightness that the peripheral retina responds to.

3.2.2 The Quantitative Description of Luminous Energy

An understanding of the mechanism by which the eye operates must depend, to a large extent, on a careful analysis of the results of a wide range of physiological photometric experiments. These are quantified in terms of the standard photometric units. Photometric units differ from most other physical units in that they are defined in terms of the subjective

sensations of vision. Thus at the very outset of a physical analysis of vision, physical science must become involved with physiology and psychology in order to frame a rational system of units. Some consideration is given in this section to the quantities and units that are used to specify and measure luminous radiation.

Physical Definitions

Fundamentally, the physical quantity involved in photometry is electromagnetic radiation, which can be described as a combination of electric and magnetic vector functions of time, $V_v(t)$ and $H(t)$, where V_v is measured in volts per meter and t is seconds. The fundamental physical reality is this combination of time-varying electric and magnetic fields. However, a much more familiar way of describing electromagnetic radiation - such as light - is in terms of the average power flux (in watts per square meter) and the relative spectral energy density. This first quantity gives the total power flux of the radiation, the latter shows how it is distributed over the frequency band. As an example, the total energy flux, E , from a black body radiator at an absolute temperature T is given by

$$E = \sigma T^4 \text{ watts}$$

where $\sigma = 5.669 \times 10^{-8}$ watts/meter²/(deg K)⁻⁴. The spectral energy density is given by the function

$$E_v = \frac{K v^3}{h v e^{\frac{h v}{k T}} - 1}$$

$$\text{where } \frac{1}{K} = \int_0^\infty \frac{v^3}{h v e^{\frac{h v}{k T}} - 1} dv = \frac{k^4 T^4 \pi^4}{h^4 15}$$

where h is Plank's and k Boltzman's constant.

This formula enables the total energy falling within a small band of frequencies to be calculated, i. e.,

$$\text{energy in band } v \text{ to } v + dv = E_v \sigma T^4 dv$$

The relationship between the ultimate physical descriptives of radiation - i. e. , the electric and magnetic vectors $V_v(t)$, $H(t)$ - and the more familiar spectral energy descriptives - i. e. , E and E_v - is through the purely mathematical procedure known as the Fourier Transform. Thus, the quantities E and E_v are not fundamental quantities. They do not, for example, completely define the functions $V_v(t)$ and $H(t)$ but on the other hand the functions $V_v(t)$ and $H(t)$ completely define E and E_v .

In the present context it seems rather artificial to describe $V_v(t)$ and $H(t)$ as fundamental quantities and E and E_v as derived quantities, since in almost all practical cases it is only the derived quantities that are of interest. This is because our sensory mechanisms (both physiological and physical) are, generally, sensitive to E and E_v . However, in certain other situations it is the Fourier Transforms that seem artificial. For example, the mathematical function that describes a two dimensional (black and white) visual scene is $B(xy)$, i. e. , brightness B as a function of position (xy) . The concept of the Fourier Transform of this function seems highly artificial, yet there is evidence to suggest that the peripheral retina may respond to the modulus of the Fourier Transform of $B(xy)$ rather than to $B(xy)$ itself. This point will be taken up later.

The mathematical relationship between $V_v(t)$ and E , is of the form:

$$E = \lim_{T \text{ Large}} \left\{ \frac{1}{2T} \int_{-T}^{+T} \alpha V^2(t) dt \right\}$$

where α is a constant related to the impedance of the medium in which the electromagnetic wave is propagated. The spectral energy distribution is given by an expression of the form;

$$E_v(v) = \alpha \int V(t) e^{-2\pi vt} dt \int V(t) e^{2\pi vt} dt$$

Photometric Definitions

In the previous section a basic physical specification ($V_v(t)$) of radiant energy was discussed and its relation to the more usual physical parameters - total energy flux E and spectral density E_v was shown. For most practical macroscopic purposes it is these latter quantities that are

of interest. The experimental basis for the energy specification of luminous radiation is that E and E_v can be measured with instruments such as the bolometer and spectroscope.

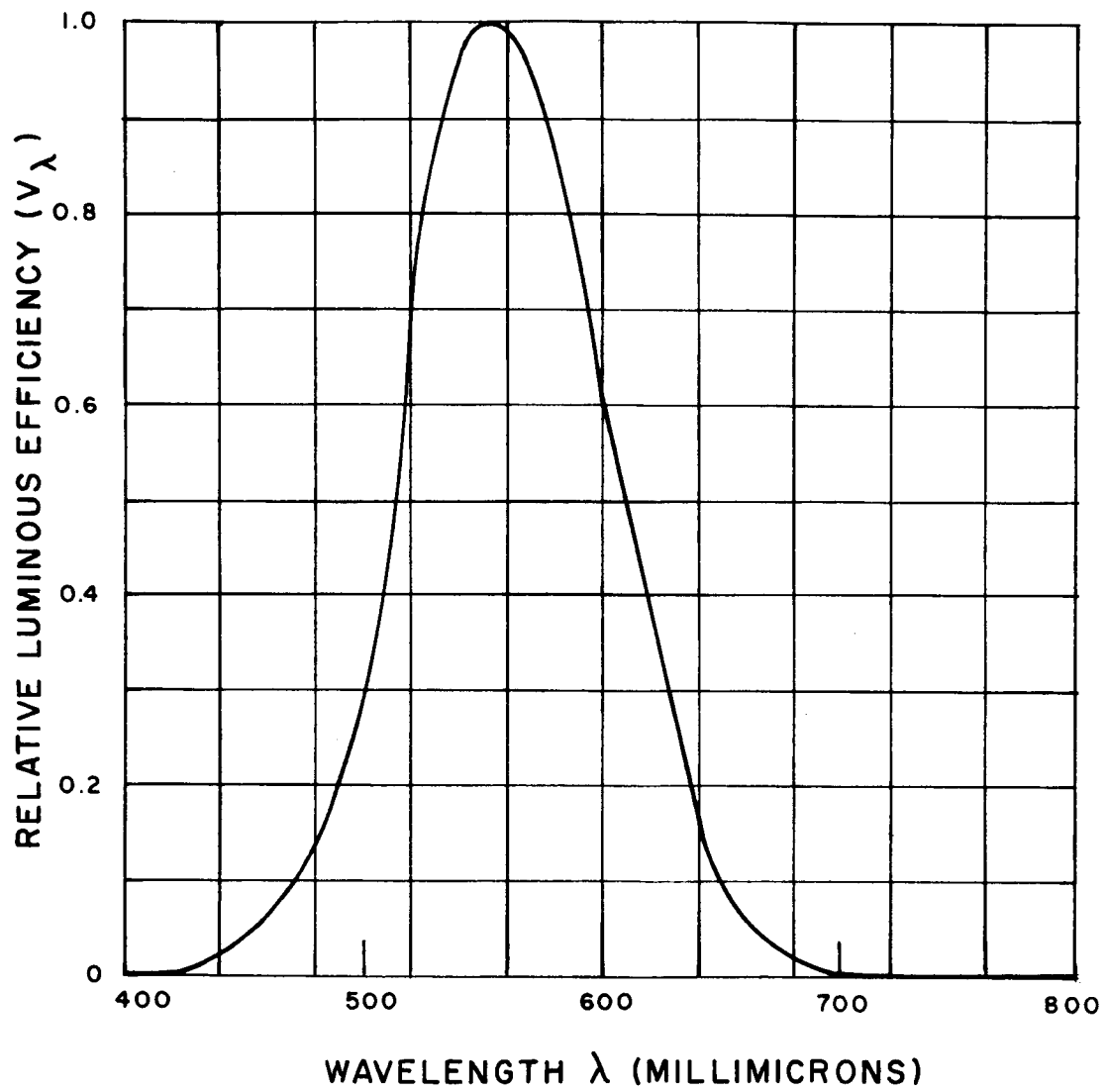
The photometric specification of luminous radiation is based on the experimental fact that human subjects can detect light and, to some extent, discriminate between differing brightness levels and colors. The photometric system of intensity is designed, as far as possible, to correspond to the subjective sensation of brightness. There are problems in doing this--principally because of the difficulty of comparing the subjective brightness of different colors. Nevertheless, to some degree at least, the subjective concept of brightness (independent of color) is meaningful and capable of measurement. It may not be possible--with great repeatable accuracy--to say at what intensity a green light is just as bright as a given red light, but it is possible to say when it is very much less bright or very much more bright. Thus the concept of subjective brightness is physically real, since experiments like this can be designed to measure it. The problem is that it cannot be measured accurately.

The fact that two light sources are judged equally bright is no indication that their energy fluxes are equal. The sensitivity of the eye to light is not constant, but is a function of wavelength. This variable sensitivity of the eye is defined by the relative luminous efficiency function, V_λ , which can be measured experimentally. Because of the Purkinje shift this function depends on the intensity of light used--corresponding to either dark or light adaptation of the retina. There are thus two luminous efficiency functions defined--the photopic (V_λ) at high intensities; and the scotopic at low intensities (V_λ^s).

These functions are subject to measurement errors and there are variations from one subject to another. Thus the concept of a standard observer is introduced, with photopic and scotopic functions calculated as the mean of the results of several investigations. The photopic function V_λ , adopted as the CIE relative luminous efficiency curve by the Commission Internationale de l'Eclairage, is plotted in Figure 3-2. The peak of this photopic curve is at 555 m μ , whereas the peak of the scotopic curve (V_λ^s) is at 507 m μ .

FIG. 3-2

C.I.E. LUMINOUS EFFICIENCY CURVE



With the adoption of this curve, photometry ceases to depend on human subjective phenomena. Photometric measurements can be made using purely physical instruments. Thus, suppose it is desired to photometrically compare two light sources defined by radiant energy functions $E_{1\lambda}(\lambda)$ and $E_{2\lambda}(\lambda)$. The photometric ratio of intensity is

$$\frac{\int E_{1\lambda}(\lambda) V(\lambda) d\lambda}{\int E_{2\lambda}(\lambda) V(\lambda) d\lambda}$$

The energy ratio of intensity is

$$\frac{\int E_{1\lambda}(\lambda) d\lambda}{\int E_{2\lambda}(\lambda) d\lambda}$$

The unit of luminous intensity in the photometric system is the candella (cd). It is such that 1 square centimeter of a black body radiator at the temperature of solidification of platinum (1773.5°C) has a luminous intensity of 60 candella. By definition, a uniform point source of 1 candella emits a luminous flux of 1 lumen per steradian. The candella and the lumen are the basic photometric units. The vast array of other photometric units relate to illumination at a surface (e.g., lumens received per unit area) and luminance of a surface (i.e., candellas per unit area). Obviously there must be some relation between the illumination of a diffusing surface and its consequent luminance. For a perfect diffusing surface this relationship can be deduced from Lambert's Law. Let a perfect diffusing surface of 1 square meter area receive an illumination of 1 lumen. This incident flux is reradiated, in all directions, by diffusion. Lambert's Law states that the flux reradiated in a direction θ to the normal is proportional to $\cos \theta$. Thus the apparent source brightness - luminance - (candles per unit projected area) is independent of the angle of view of the surface. Let this luminance be I candles per square meter. The flux radiated at an angle θ to the radiating surface through an elemental

surface area of a unit hemisphere is

$$\begin{array}{lcl} dF = I \cos \theta & \times & 2\pi \sin \theta d\theta \text{ Lumens} \\ \text{(source intensity)} & \times & \text{(solid angle)} \end{array}$$

The total flux radiated from the surface is thus:

$$F = 2\pi I \int_0^{\pi/2} \cos \theta \sin \theta d\theta = \pi I$$

But F was set at 1 lumen. Therefore

$$\begin{array}{l} \pi I = 1 \\ I = \frac{1}{\pi} \text{ candles per sq.} \end{array}$$

Thus when the illumination is 1 lumen per square meter, the illuminance (of a perfect surface) is I/π candles per square meter. When the illumination is 1 lumen per square centimeter (1 phot) the luminance of the (perfect) surface is said to be 1 lambert (or equivalent phot). Thus 1 lambert corresponds to a luminance of $\frac{1}{\pi}$ candles per square centimeter. Other photometric units are shown in Table I.

Mechanical Equivalent of Light

The relationship between the photometric system (lumens) and the physical system (watts) can be determined. If the radiant energy through a surface is defined by the function $E_\lambda(\lambda)$ (watts per square meter), the luminous flux (in lumens per square meter) is given by

$$L = Km \int E_\lambda(\lambda) V_\lambda(\lambda) d\lambda$$

where the integral extends, theoretically, over the whole electromagnetic spectrum.

Since the photometric units are based on a quite arbitrary standard source it is natural that an arbitrary constant, Km, must be introduced in the above equation. Its value can be calculated by taking for E_λ the known energy distribution curve for 1 square cm of a black body at 1773.5°C . The function $V_\lambda(\lambda)$ is the photopic luminous efficiency function (Fig. 3-2). L must be set at 60 since this is the defined intensity

TABLE I
PHOTOMETRIC UNITS

<u>Source</u>	candella (cd) (new candle). Defined in terms of freezing platinum.	
<u>Flux</u>	lumen Defined as the flux radiated into 1 steradian by a uniform point source of 1 candella.	
<u>Illumination</u>	Flux (lumens) per unit area incident of surface	
Metre-candle		1 lumen/sq. metre
Lux		1 lumen/sq. metre
Phot		1 lumen/sq. cm
foot-candle		1 lumen/sq. ft.
<u>Luminance</u>	Source (candellas) per unit area of surface	
1 stilb		1 cd/sq. cm
1 nit		1 cd/sq. in
1 equivalent phot		3183 cd/sq. in
1 lambert		3183 cd/sq. in
1 equivalent lux		0.3183 cd/sq. in
1 blandel		0.3183 cd/sq. in
1 apostilb		0.3183 cd/sq. in
1 equivalent foot candle		0.3183 cd/sq. foot
1 foot lambert		0.3183 cd/sq. foot

TABLE I (Con't)

<u>Luminance</u>	resulting from	<u>Illumination</u>
1 candella per sq. cm		II lumens/sq. cm
1 stilb		II lumens/sq. cm
1 nit		II lumens/sq. in
1 candella per sq. ft.		II lumens/sq. ft.
1 equivalent phot		1 lumen/sq. cm
1 lambert		1 lumen/sq. cm
1 equivalent lux		1 lumen/sq. in
1 blandel		1 lumen/sq. in
1 apostilb		1 lumen/sq. in
1 equivalent foot candle		1 lumen/sq. ft.
1 foot lambert		1 lumen/sq. ft.

The above Table is only correct when the surface being illuminated is a perfect diffuser.

(in candella) of the specified source. The result of this calculation is quoted in Reference 3-2 as $K_m = 679.6$ lumens/watt and in Reference 3-3 as $K_m = 678.8$ lumens/watt. In this report the first value will be assumed.

If instead for the photopic curve $V_\lambda(\lambda)$, the scotopic function $V'_\lambda(\lambda)$ is substituted the result is $K_m = 1745.0$ lumens/watt (Ref. 3-2).

K_m is sometimes termed the mechanical equivalent of light. Thus in the photopic system, 679.6 lumens of light at the wavelength corresponding to the peak of the photopic luminous efficiency curve ($\lambda = 555 \text{ m}\mu$) corresponds to one watt.

The relation between photometric and energy units will usually involve an integration process. Thus let $E_\lambda(\lambda)$ define the spectral energy density of a source. The total power output of the sources is thus

$$W = \int_0^\infty E_\lambda(\lambda) d\lambda \text{ watts}$$

The luminous intensity of the sources is

$$L = K_m \int E_\lambda(\lambda) V_\lambda(\lambda) d\lambda \text{ lumens}$$

$$\therefore L = \frac{K_m \int E_\lambda(\lambda) V_\lambda(\lambda) d\lambda}{\int E_\lambda(\lambda) d\lambda} W$$

Quantum Theory

The sensation of vision depends, essentially, on the absorption of radiation. Thus the quantum nature of light is of some significance. This provides yet another quantitative method of describing light energy--namely in terms of the number of quanta per second passing through unit area.

At $\lambda = 555 \text{ m}\mu$, a flux of N quanta/sec. corresponds to an energy flux of

$$\begin{aligned}
 W &= N h \nu = \frac{N h c}{\lambda} \quad \text{watts} \\
 &= N 6.625 \times 10^{-34} \frac{3 \times 10^8}{5.55 \times 10^{-7}} \quad \text{watts} \\
 &= 3.58 \times 10^{-19} N \quad \text{watts}
 \end{aligned}$$

At this wavelength, 1 watt = 679.6 lumens. Therefore 1 lumen
 $= \frac{10^{19}}{3.58 \times 679.6} = 4.11 \times 10^{15}$ quanta/sec. An interesting, and pertinent,
 calculation is the quantum flux in an image formed by a lens.

Let the object be a plane surface, in an illumination of
 I lumens/sq. m, at a distance x from a lens of focal f and aperture D
 (meters).

Consider an element of area of the object, that is normal
 to the lens, and which subtends 1 square second of arc at the lens. Its
 area will thus be

$$\left(\frac{1}{3600} \frac{\pi}{180} \right)^2 x^2 \quad \text{square meters.}$$

The luminance of the surface is I/π candles per square
 meter (assuming it to be a perfect diffuser). Thus this small element of
 area subtaining 1 sq. second at the lens will act as a source of intensity

$$\frac{I}{\pi} \left(\frac{1}{3600} \frac{\pi}{180} \right)^2 x^2 \quad \text{candellas.}$$

The fraction of the light emitted by this source that will reach the lens is

$$\begin{aligned}
 &= \frac{\pi D^2}{4x^2} \left(\frac{1}{3600} \frac{\pi}{180} \right)^2 \frac{I}{\pi} x^2 \quad \text{lumens} \\
 &= \frac{\pi D^2}{4} \left(\frac{1}{3600} \frac{\pi}{180} \right)^2 4.11 \times 10^{15} \quad \text{quanta/sec} \\
 &= 2.41 \times 10^4 I D^2 \quad \text{quanta/sec. /sq. second of arc}
 \end{aligned}$$

In the case of the human eye, we may set $D=5$ mm for an average pupil diameter and take $I=200$ lumens per sq. m (for an average room illumination). Then the number of quanta involved is

$$\begin{aligned} & 2.41 \times 200 \times 25 \times 10^{-6} \times 10^4 \\ & = 1.2 \times 10^2 \text{ quanta/second/sq. second of arc.} \end{aligned}$$

This calculation ignores the losses in the eye. It may be assumed that only about 10% of the quanta incident at the eye are actually absorbed by the photoreceptors. At an illumination of 200 lumens/sq. m the summation time of the eye is of the order of a 1/30 second. Thus the number of quanta absorbed per summation time per square second of arc is only 0.4.

This order of magnitude calculation is of interest in showing that detail finer than about one second of arc simply does not exist in the retinal image--independently of any other limitations that may exist such as diffraction, blurring, etc.

Retinal Units

The illumination of the retina depends on the brightness of the viewed scene and also on the size of the pupil. Since in many visual experiments the quantity of interest is retinal illumination, a unit has been defined incorporating both illumination and pupil size. This unit is the troland (formerly called photon). An extended source of white light of luminance 1 candella/sq. m seen through a pupil of 1 sq. mm area aperture is defined to yield a retinal illumination of 1 troland.

When the scene illumination is I lumens/sq. m the luminance is I/π candella per sq. meter. If the pupil diameter is D meters, the area is $\frac{\pi D^2}{4}$ sq. meters. Thus the retinal illumination T (in trolands) is

$$T = I/\pi \times 10^6 \frac{\pi D^2}{4} = ID^2 \times 2.5 \times 10^5 \text{ trolands}$$

When D is expressed in millimeters $T = 0.25 ID^2$ trolands. D as a function of I is plotted in Figure 3-3 (an approximate average of several curves given in Reference 3-3). The corresponding function of T against I is plotted in Figure 3-4, which must be regarded as an approximation only.

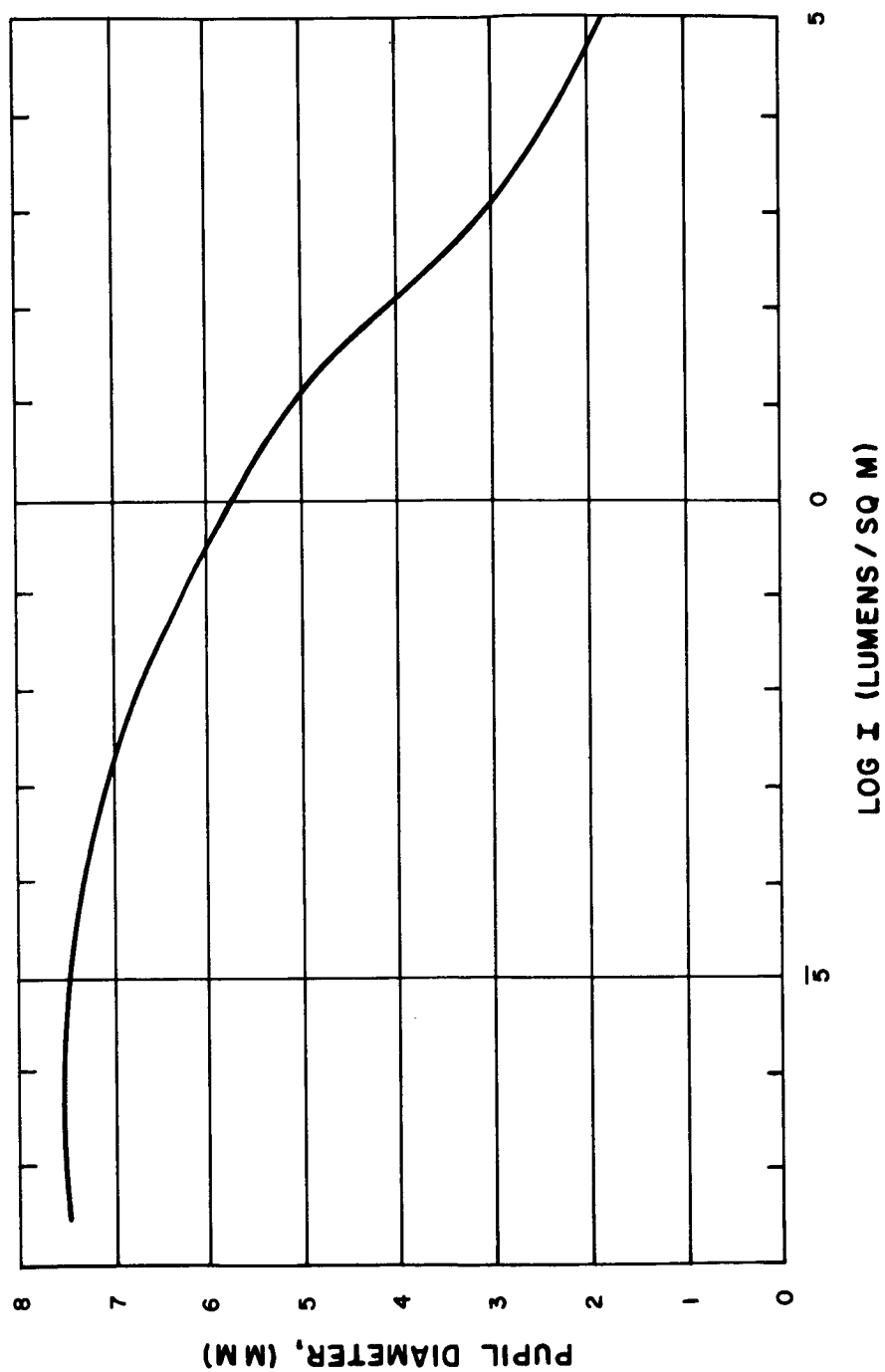
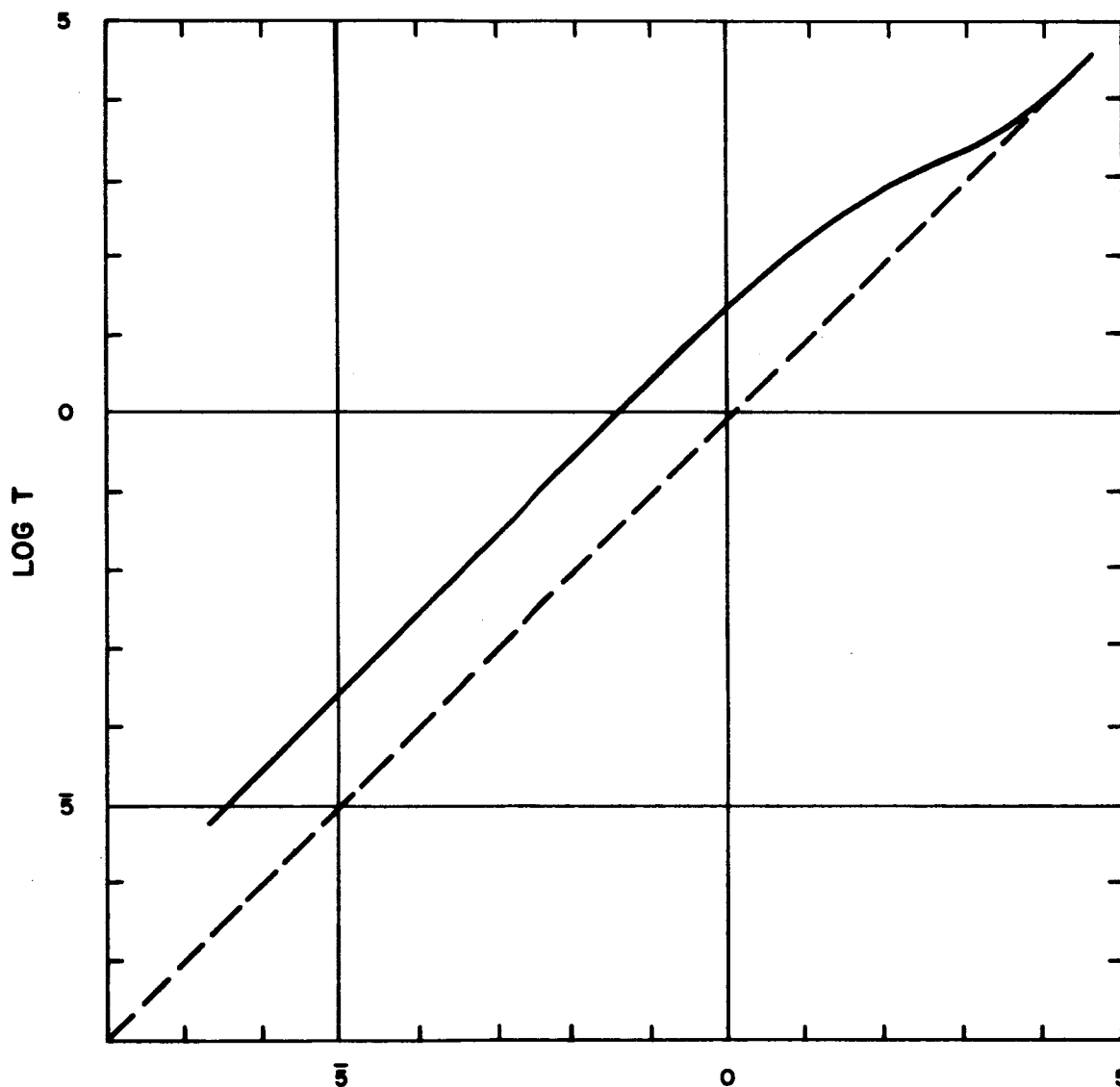


FIG. 3.3
PUPIL DIAMETER AS A FUNCTION
OF RETINAL ILLUMINATION (APPROXIMATE)

FIG. 3-4
APPROXIMATE RELATIONSHIP BETWEEN
TROLLANDS AND LUMENS/SQ M
(BASED ON FIG. 3-3)



3.2.3 The Image Forming Process

Optical System

The essential features of the eye, considered as an optical instrument are:

1. the refracting system; which includes a lens of variable refractive power,
2. a photosensitive screen on which is formed an image of the scene in front of the eye,
3. a pupil, the diameter of which is controlled by a sub-conscious reflex system.

The refractive properties of the eye can be define explicitly or in the form of various semiequivalent schematic eyes (Ref. 3-4). The refractive power of the eye is expressed in dioptries--the reciprocal of the distance in meters of the fixation point from the eye.

Transmission Characteristics of the Ocular System

The spectral transmittance of the eye from the cornea to the retina, as measured by Ludvigh and McCarthy (Ref. 3-5) is shown in Figure 3-5. For foveal vision the transmittance of the macula pigment is of significance and must be allowed for.

The short wavelength cut-off of the eye appears to be due to the transmission of the ocular media. An eye from which the lens has been removed is sensitive to shorter wavelengths than a normal eye. The retina can be stimulated by very short electromagnetic rays, i. e., Xrays.

It can be seen from Figure 3-5 that the transmission of the ocular media is about 0.5 in the middle of the visible spectrum. The fraction of the light incident at the cornea that is actually absorbed by the retina is important for quantum effect calculations. In Reference 3-2 it is estimated that 30% of the light falling upon the rods is absorbed. Allow-
ing for the fact that the retina contains cones and interspaces it is calculated that only 20% of the light incident upon the retina is absorbed by the rods. Thus, according to this calculation, the total efficiency of the rod system is only $0.5 \times 20\% = 10\%$.

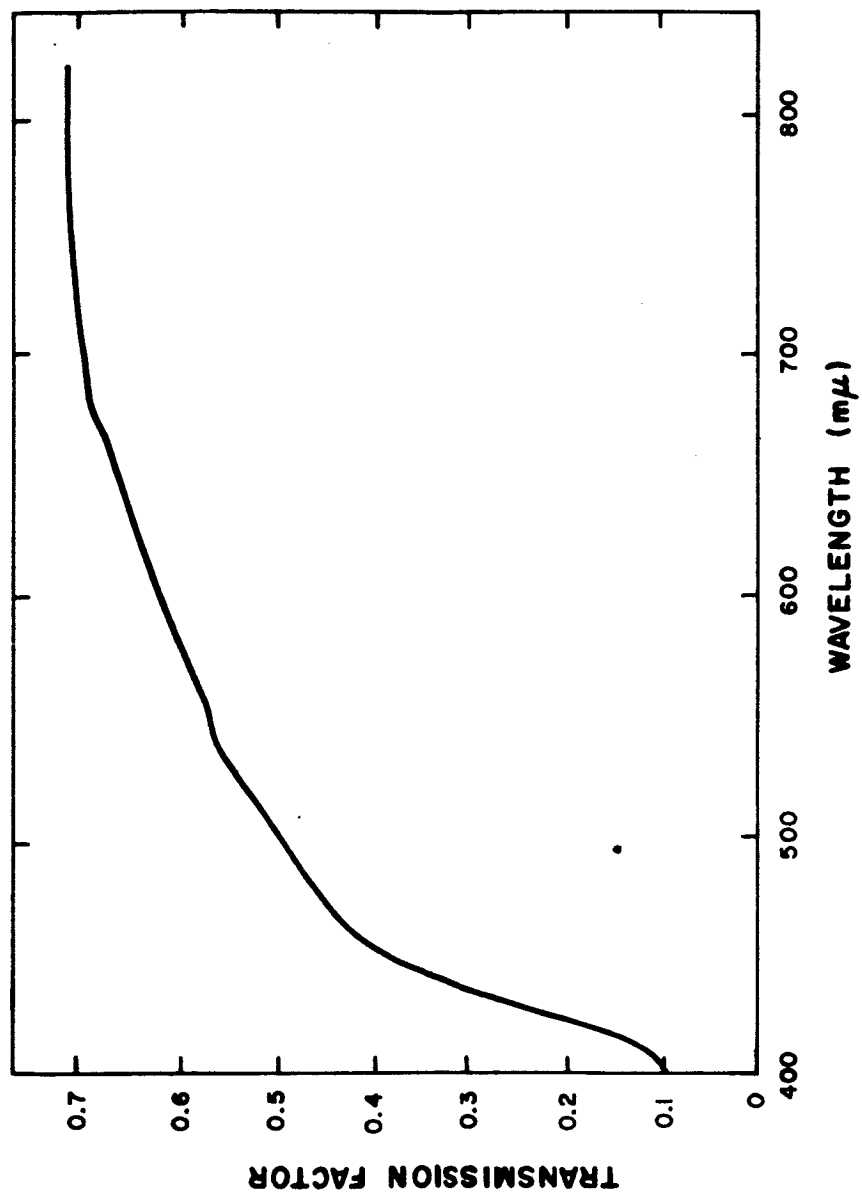


FIG. 3-5
TRANSMISSION FACTOR
OF THE HUMAN EYE (REF 3-5)

Defects of the Image

It must be emphasized that the retinal screen covers a very large angular area--corresponding to the large receptive field of the static eye. The equivalent lens system of the eye, and the sharpness of the image produced, etc., are generally described, however, only for the very small foveal point of the retina. No data has been found that describes the quality of the peripheral image, although there is little doubt that it is relatively poor.

The principle defects of the retinal image are:

1. blur due to;
 - (a) incorrectly focused image
 - (b) astigmatism
 - (c) diffusion within the eye
2. chromatic aberration
3. diffraction

Defects caused by abnormalities will be ignored. For the present, chromatic aberration will be considered as a form of defocusing.

It is clear that the accuracy of focusing cannot be greater than the sensory resolving power of the retina. The focusing mechanism is a closed loop servo in which the retina is the sensor, and thus it would not be possible for the focusing action to be more precise than the sensing capabilities of the retina. Correspondingly, it is extremely unlikely that the focusing mechanism could be significantly less precise than retinal acuity, for then the retina would have a sensory capability it could never utilize. It is concluded, therefore, that the focusing mechanism is probably not an independent limitation of, at least, the foveal image.

For small visual fields, according to Flamant (Ref. 3-6), diffusion scatters the light over a larger area than that caused by diffraction. The spatial distribution, by diffusion, of light from a point source is, however, of the same general shape as the diffraction pattern up to the first zero.

Image distortion caused by diffraction can be defined in terms of the brightness pattern formed on the retina by the image of a well focused point source of light.

Diffraction phenomena are classed as either Fraunhofer or Fresnel. Fraunhofer diffraction occurs when source and image are both effectively at infinity. This is approximately the case in the eye.

The use of sine wave analysis in optics is discussed in Section 3.2.6. It is shown, for example, that the electromagnetic wave intensity on a screen due to Fraunhofer diffraction of a point source is The Fourier Transform of the function that defines the transmission of the aperture. The brightness of the diffraction pattern is the square of the electromagnetic wave intensity and thus can be described as the square of the modulus of the Fourier Transform of the aperture function. The relationship between the diffracting aperture and the diffraction pattern is entirely analogous to that between an electronic network (e.g., a low pass filter) and the square of its impulse response. Thus, just as the network can be described as a filter so also the diffracting aperture can be described as a spatial filter. The electronic filter limits the high frequency components in the electronic signal, the aperture limits the high frequency, spatial components in the image. It can be shown that a circular aperture of diameter D sets an upper limit to the frequency variable $\Omega = \frac{2\pi D}{\lambda}$, i.e., a frequency of D/λ lines per radian in the diffracted image.

Image defects due to incorrect focusing can be assessed by assuming that the focal length of the eye is in error by δ . Then what should be a point image will yield a circle of confusion of diameter $\frac{\delta}{f} D$, where f is the focal length of the eye and D is the diameter of the pupil. Sample points in the image must be then separated by $\frac{\delta D}{f^2}$ radians. This corresponds to a cut off frequency of $\frac{f^2}{2\delta D}$ lines/radian.

The Quantum Nature of the Visual Process

From the point of view of its ultimate photoabsorption by the retina, the retinal image can be considered as a hail of quanta. The discrete nature of this flux limits the resolution obtainable--independently of diffraction and defocusing effects and also of the sensory characteristics of the retina. The quantum nature of light is thus of great significance in this study of the information sensing characteristics of the retina.

Suppose that the eye is viewing a scene with an illumination of I lumens/sq. m (i.e., I lux). From the calculation given in Section 3.2.2,

(under "Quantum Theory") the quantum flux at the retina will be $2.41 \times 10^4 ID^2$ quanta/sec./sq. second of arc, where D is the pupil diameter in meters. Converting the formula so that D is in millimeters and the unit of area is 1 square degree, the flux is $3.14 \times 10^5 ID^2$ quanta/sec./square degree, where D is now in millimeters. To account for transmission losses in the eye and in the retinal absorption process a factor, K, must be introduced which will have a value of approximately 0.1. Furthermore the retina integrates the received flux for a period of time τ , i. e., the summation time. Thus the actual quantum flux that is available to the nervous system for interpretation as a visual sensation is

$$3.14 \times 10^5 ID^2 K \tau \text{ quanta/square degree}$$

This result can be expressed in trolands, T (section 3.2.2, under "Retinal Units"). That is the quantum flux

$$\begin{aligned} &0.25 \times 3.14 \times 10^5 TK\tau \\ &= 0.78 \times 10^5 TK\tau \text{ quanta/square degree/per summation} \end{aligned}$$

time. (where T is in trolands)

The summation time can be taken as the reciprocal of the critical fusion frequency. The value of this flux (assuming $K = 0.1$) over a wide range of illumination is plotted in Figure 3-6. It can be seen that at the absolute threshold the quantum flux is so low that only objects as large as about 5° will yield even 1 quanta per summation time. It is obvious, therefore, that at the absolute threshold only a very vague outline of the scene will be observed. At the other end of the range--10 decades higher illumination--it is clear that the visual process is still quantum limited. For example, in ordinary office illumination the effective quantum flux is approximately 1 quanta per 20 sq. seconds. Detail of the order of 1 sq. minute represents the limit of detail that can be seen. For very fine work a higher illumination still is recommended (for example in a drafting office). Thus over most of the operating range of illumination of the eye the visual process is quantum limited. Information is extracted by the (foveal) retina up to the limit set by the discrete nature of the quantum flux.

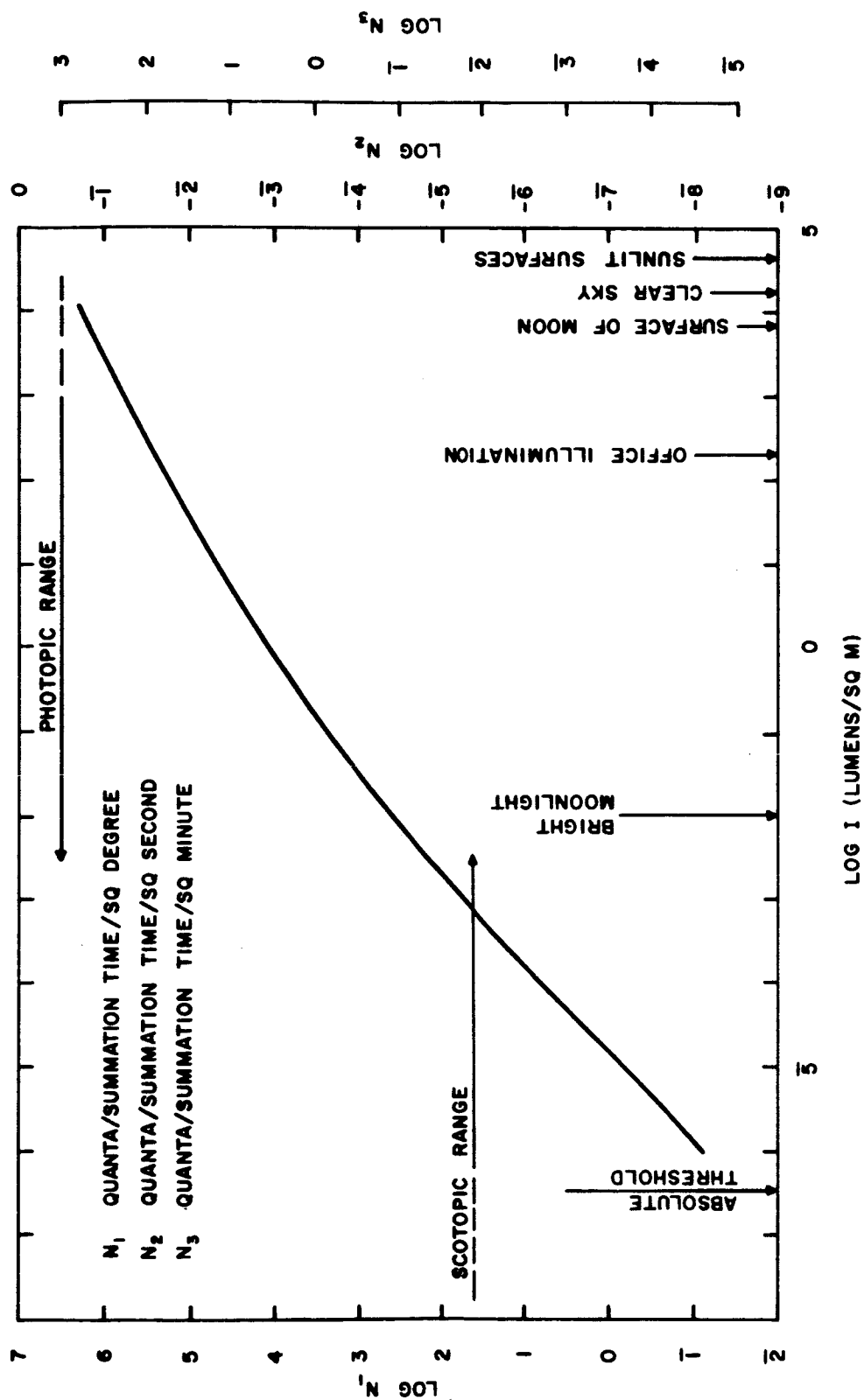


FIG. 3-6
QUANTUM FLUX

Definition of Input Image

Three important defects, or limitations, of the retinal image have been considered:

1. defocusing
2. diffraction
3. quantum limitations

Each of these effects can be described as a limitation of the spatial bandwidth (in lines per degree) of the retinal image. Thus even before sensory detection, the information content ('bandwidth') of the retinal image is limited.

In the case of quantum limitations, a sampling interval must be chosen. Suppose that the average flux is 3600 quanta per square degree. If the sampling points are taken at degree intervals very accurate brightness information will result, since the statistical fluctuations in a number as large as 3600 will be of the order of 60, i. e. about 1.6%. If the sampling intervals are taken every minute of arc the average number of quanta per sample will be only 1, and thus the statistical fluctuations will amount to 100%. In the first case, very accurate, low frequency information is obtained. In the second case very inaccurate, high frequency data results. There is obviously an optimum sampling interval.

Let there be an average flux of N quanta per square degree. Let the sampling interval be θ degrees. Then each sample will contain $N\theta^2$ quanta. There will be $1/\theta^2$ samples per square degree. Due to statistical fluctuations the sample $N\theta^2$ does not give exact brightness information. There is an RMS error of $\theta\sqrt{N}$ associated with such a sample. Regarding $N\theta^2$ as "noise" + "signal", and $\theta\sqrt{N}$ as "noise", Shannon's formula can be applied to compute the information yielded per square degree:

$$\begin{aligned} &= \frac{1}{\theta^2} \log_e \frac{N\theta^2}{\theta\sqrt{N}} \text{ bits} \\ &= \frac{1}{\theta^2} \log_e \theta \sqrt{N} \end{aligned}$$

θ may be chosen to maximize this quantity: setting the differential to zero

$$2/\theta^3 \log_e \theta \sqrt{N} = \frac{1}{\theta^2} \frac{\sqrt{N}}{\theta \sqrt{N}}$$

thus

$$\log_e \theta \sqrt{N} = \frac{1}{2}$$

$$\theta^2 N = e$$

The "bandwidth" f ; in lines per degree, is given by

$$f = \frac{1}{2\theta} = \frac{\sqrt{N}}{2\sqrt{e}}$$

f , as a function of I is plotted in Figure 3-7 from the results shown in Figure 3-6. The diffraction limit on image bandwidth was shown in Section 3.2.3 (under "Defects of the Image") to be a frequency of D/λ lines/radian where D is the pupil diameter and λ the wavelength. Taking $\lambda = 555 \text{ m}\mu$ this frequency, f , is given by

$$f = \frac{D}{5} 10^4 \text{ lines/radian where } D \text{ is now in millimeters}$$

$$= \frac{\pi D}{5 (180)} 10^4 \text{ lines/degree}$$

$$= 35 D \text{ lines/degree}$$

The bandwidth limitation due to defocusing is given in Section 3.2.3 as

$$f = \frac{F^2}{2\delta D} \text{ lines/radian}$$

where F is the (internal) focal length of the eye, D the pupil diameter and δ the error in the focal length. Changing the units

$$f = \frac{17^2 \pi}{2\delta D (180) 10^{-3}} \text{ lines/degree}$$

$$= \frac{2.52 \times 10^3}{\delta D} \text{ lines/degree}$$

where D is in mm and δ in microns.

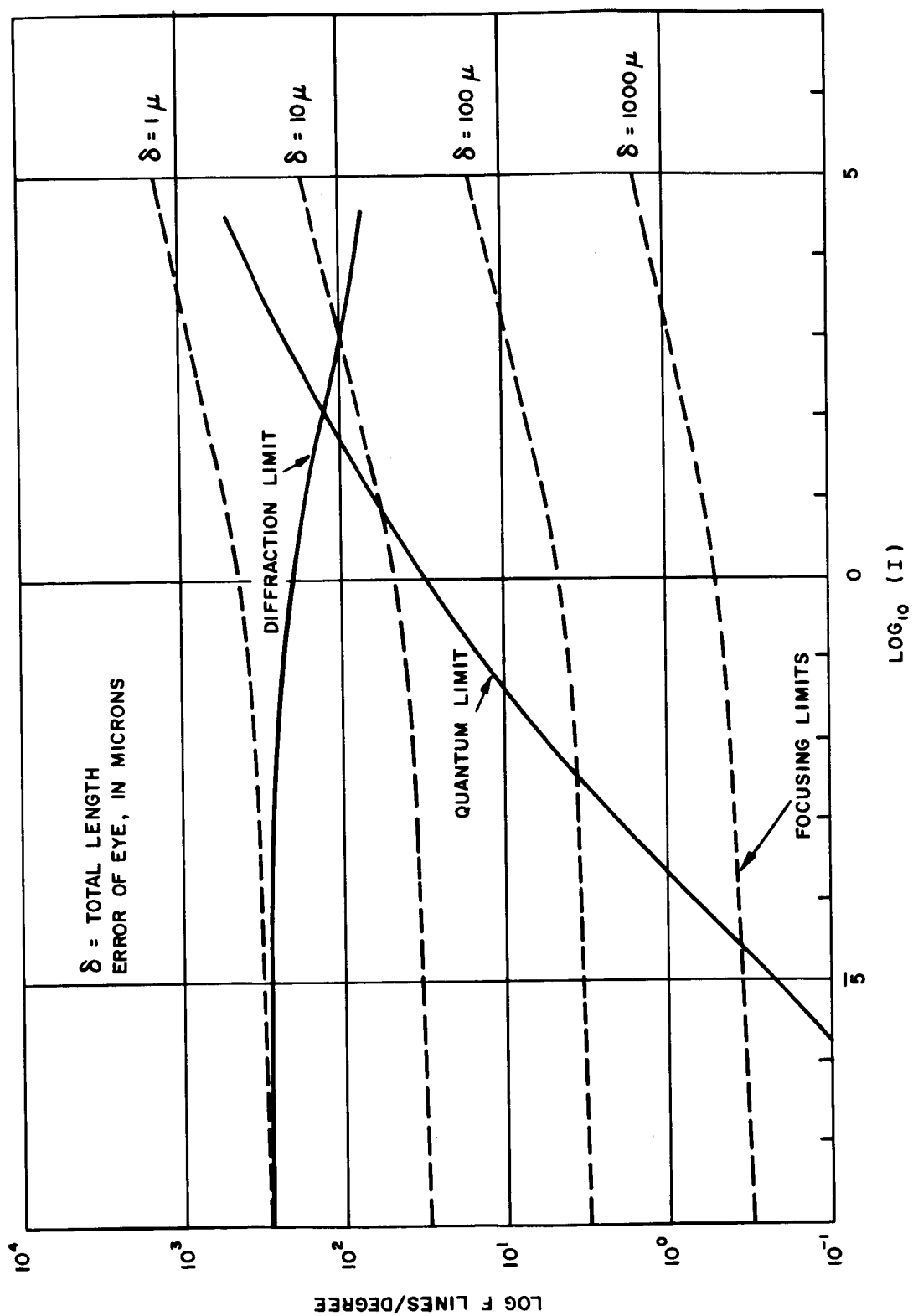


FIG. 3.7
BAND WIDTH OF THE EYE

3.2.4 Image Analysis

The human visual system has been analyzed as an information sensor. It has been seen how the focusing mechanism, pupil diffraction, and the quantum nature of light energy each set limits to the spatial frequency bandwidth of the eye -- and thus to its ability to extract information from a viewed scene.

An important aspect in the analysis of any signal transmission system is the nature of the signals themselves. For example in television the general nature of the signal can be defined in terms of its mean power spectral density. This will show that while frequency components exist up to several megacycles, most of the energy is at low frequencies, and the energy is also spaced at regular intervals throughout the frequency spectrum -- corresponding to the harmonics of the line and frame frequencies. Similarly the average speech signal on telephone circuits can be defined in terms of a power energy density function.

An immediately significant feature of images -- thought of as signals -- is that they are two dimensional, and that they can contain two general types of 'elements' -- namely points and lines. Suppose the image is in the form of a square and contains n^2 individual resolvable point elements. It is possible to compute how many different straight lines can be drawn on this square. (The various segments of one straight line will not be considered as different lines). Obviously n^2 lines can be drawn crossing two given edges of the square. Since there are 6 ways of choosing two edges there are $6n^2$ lines altogether. The points within the square may not always exactly coincide with these lines but nevertheless can be associated with each line to define it. Each of these lines will consist of approximately n points. Thus a "message" consisting of n points in the form of a straight line may have $6n^2$ different 'values'. However, "messages", consisting of n points chosen at will, may assume approximately $(n^2)^n$ different "values". It is clear therefore that by restricting an image so that it is composed entirely of straight lines imposes a very large restriction on its information content.

Such general restrictions are often imposed in the case of (one dimensional) radio signals. For example the telemetry systems of planetary probes are designed to operate at very low bit rates -- of the order of 1- 100 bits/second. In this particular example the transmitter power is inevitably limited and the received signals must be picked out of the thermal noise background.

For a given transmitter power, the signal to thermal noise ratio improves as the bit rate is reduced.

Analogously, in vision, a signal -- ie. an image -- consisting of straight lines, stands out vividly against the 'noise' which limits the performance of the visual sense. This sort of argument can obviously be extended to cover curved lines and segments of lines, etc. The essential point is that these elements carry relatively little information compared to the same number of points distributed at random, and correspondingly are easier to discriminate.

The conclusion is that the fact that the eye seems to sense line contours in an image -- as evidenced by the ability of an artist to catch a likeness of a face with just a few line strokes of a pencil -- can be ascribed to the higher signal/noise ratio of such patterns.

It is possible to point out two general types of applications of this feature of vision;

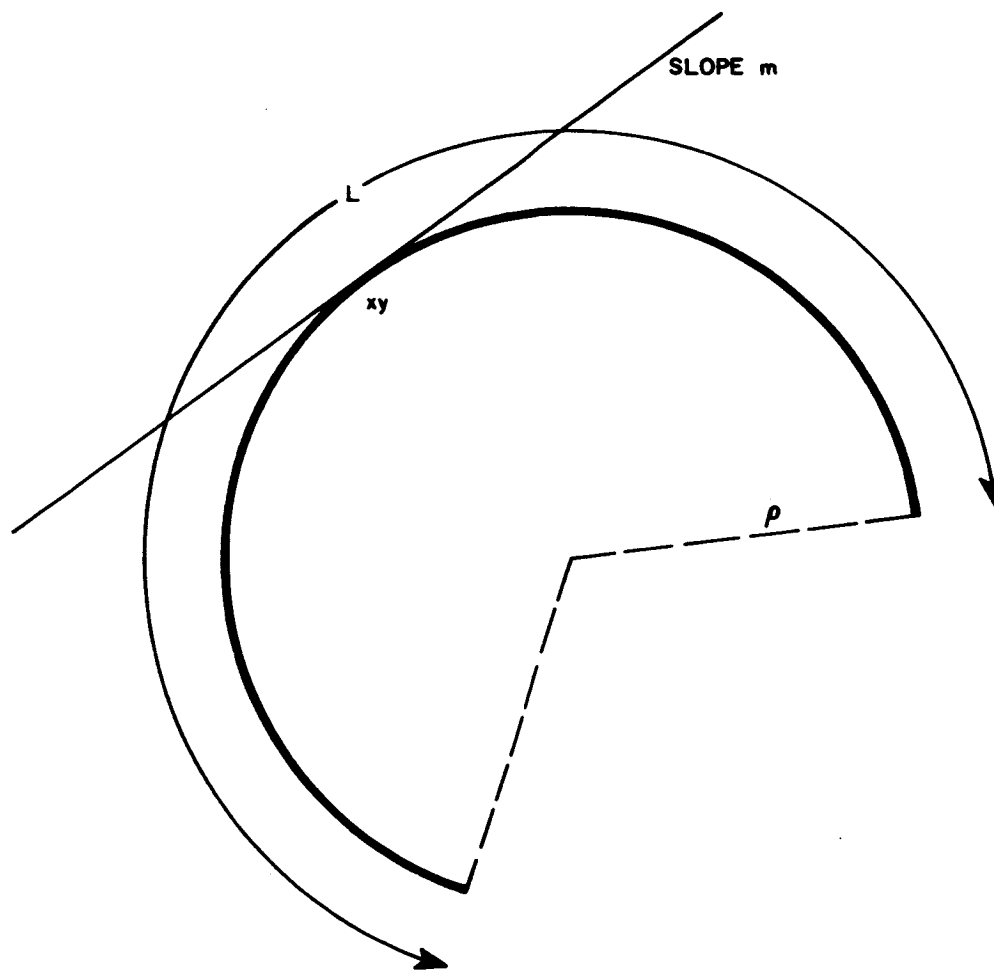
(a) in television, or any optical system, that may be interposed between a man and the environment he wishes to visually observe.

(b) in a completely automatic system, in which visual sensing of an environment is involved.

In conventional television no special attempt is made to utilize the bandwidth saving properties of line images. As an example of the savings that might be effected, consider the special problem of transmitting the images of alphabetic characters. To produce clear, sharp images a square mosaic of at least 100 elements (i.e., 10×10) is needed. That is, the "message" will be 100 brightness values, b_i where $i = 1, 2, \dots, 100$. An alternative method would be to transmit the length (L), the radius of curvature (ρ), the coordinates of the center (xy) and the slope (m) at (xy) of each segment (see Fig.3-8), of a character (see Fig.3-9). It can be seen that this second method requires an average of only about 15 numerical values per character as against at least 100 by the straightforward matrix method.

The application to completely automatic systems of those principles of image analysis that the eye apparently exploits has great potential. Generally speaking inanimate optical sensing devices (ie with no human element in the system) are extremely crude, eg. a photocell to control the opening and closing of a door, the IR sensor in the Sidewinder missile, etc. These may be described as "one-bit" devices - that is they can sense only the presence or absence of radiant energy.

FIG. 3.8
SEGMENT COORDINATES



CHARACTER	NO. OF SEGMENTS	TOTAL NO. OF COORDINATES
A	4	20
B	3	15
C	1	5
D	2	10
E	4	20
F	3	15
G	2	10
H	3	15
I	3	15
J	3	15
K	3	15
L	2	10
M	4	20
N	3	15
O	1	5
P	2	10
Q	2	10
R	3	15
S	2	10
T	2	10
U	3	15
V	2	10
W	4	20
X	2	10
Y	3	15
Z	3	15
AVERAGE NO. OF COORDINATES/CHARACTER		<u>13.7</u>

Fig. 3-9. Segment Coordinates of Alphabetic Characters

Almost all biological organisms -- above a certain primitive level -- employ optical sensing mechanisms that operate, not with a simple yes-no feature of the visual scene, but with complex patterns. For example, a bird can find its way back to its nest by visual sensing of the pattern of its environment. The duplication in inanimate systems of this general sort of capability is probably in the future, for it may depend on developing data processing systems comparable to the brain systems of nature. However it is very probable that special optical sensors for these systems will have to be developed to supply meaningful information about the visual scene to such a central data processing unit. Television might, for example, not be adequate since it slavishly reports the point by point brightness of the scene. More appropriate perhaps would be a device that would sense the segment parameters (L, ρ, x, y, m) discussed earlier. The total amount of information in a 2π steradian view may be so great that if this latter approach is not followed the sensing part of the total system may become infeasible. Thus the human eye can, under certain conditions, resolve angles as small as 1 second of arc. If this large angular area (2π steradians) were to be sensed to this resolution (1 second of arc) by a conventional TV system it would have to handle more than 10^{11} resolvable elements per frame. This is more, even, than the total number of brain cells.

Going one step further, the development of machines to read typewritten or hand written script may be considered. In this case, it is not simply a question of detecting line segments and contours but of associating the alphabet with various patterns. What characteristic of a pattern determines that it is, in fact, a letter 'a'? With hand written script a very large, if not infinite, number of different patterns must be interpreted as 'a'. Ultimately the only conclusive factor in making this determination is that all, or almost all, educated people can recognize it as 'a'. In other words the essential characteristics that distinguish a pattern in this way is to be found only in the structure of the (educated) brain. This implies that very high performance script reading machines will have to include a data processing system comparable to this capability of the human brain.

3.2.5 The Application of Quantum Considerations to Visual Sensing

Introduction

In the case of vision, the quantum nature of light energy sets the ultimate limit to the amount of information contained in an image. The quantum efficiency of the eye is high -- of the order of 10%. It may reasonably be postulated that an organism is sensitive to all the information contained in the detected signal. It follows, therefore, that many details of visual sensing should be relatable to the quantum information limit of the detected signal.

Color Vision

Even though the corresponding three different photoabsorbing pigments have not been identified, the tri-stimulus theory of color vision would seem to be unassailable, on the basis of the results of color matching experiments.

"...the reduction to 3 of the infinity of variables required to describe the phenomena of visible radiation in physical terms, clearly indicates the intervention of simplifying mechanisms of biological origin." (Ref. 3-3)

It has been suggested that the actual tri-stimulus mechanism may not involve different photopigments but may be related to standing wave phenomena in the cones, acting as optical wave guides.

The actual mechanism is of no interest in the present context. It will simply be assumed that the amount of energy falling under each of the three spectral curves is determined. Thus, it is assumed that there are three separate light sensing mechanisms in the retina, each with its own spectral distribution curve $S_i(\lambda)$ ($i = 1, 2, 3$). The magnitude, S_i , is defined as the ratio of the number of quanta incident at the retina, to the number effectively detected, per summation area per summation time. (Fig. 3-10).

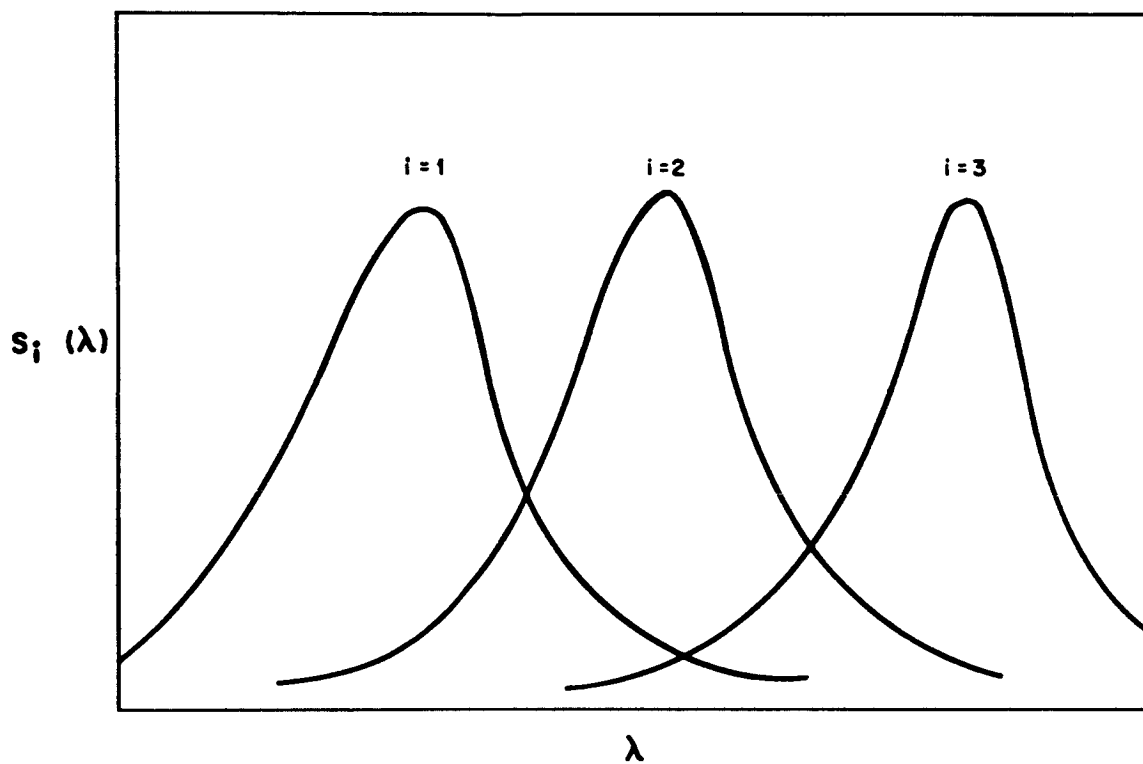
The basic experimental fact leading to the tri-stimulus theory is that any light at all can be matched in terms of subjective effect by a suitable mixture of white light and a single monochromatic light, ie.

$$\Sigma L(\lambda_\tau) = L_\omega + L_\lambda \quad (3-1)$$

where $L(\lambda_\tau)$ is the luminance of light at wavelength λ_τ ($\tau = 1, 2, 3, \dots$) (i. e. the left hand side represents a general light source) and L_ω the appropriate white light luminance and L_λ the appropriate luminance of monochromatic light of

FIG. 3.10

SPECTRAL CURVES



wavelength λ . The purity p is defined as

$$p = L_{\lambda} / L_{\omega} \quad (3-2)$$

and λ is termed the dominant wavelength. The total luminance is defined as

$$L = L_{\omega} + L_{\lambda} \quad (3-3)$$

The three parameters L , p , and λ can define -- in terms of subjective experience -- any light source at all. Let U_i be the number of quanta per summation time, per summation area, detected by the i^{th} mechanism ($i=1, 2, 3$) when light defined by L , p , λ is incident at the eye.

U_i can be considered as made up of two components -- that due to the monochromatic light, and that due to the white light. The white light luminance is

$$L - pL = L(1-p)$$

and the chromatic luminance is pL

Let

$$\sigma_i = \frac{1}{\lambda_m - \lambda_o} \int_{\lambda_o}^{\lambda_m} S_i(\lambda) d\lambda \quad (3-4)$$

(where λ_m , λ_o define the upper and lower limits of the visible spectrum)

Then the contribution to U_i of the white light component is

$$U_i = \sigma_i L(1-p)$$

and that due to the chromatic component is

$$U_i = S_i(\lambda) pL$$

Thus altogether

$$U_i = \sigma_i L(1-p) + S_i(\lambda) pL$$

Let

$$M_i = S_i - \sigma_i$$

Then

$$U_i = \sigma_i L + pLM_i \quad (3-5)$$

The information received by the brain is the magnitude of the three quantities U_1, U_2, U_3 . With each of these quantities there will be associated a shot noise with an RMS amplitude of $\sqrt{U_i}$. This shot noise will determine how small a change in the stimulus may be reliably identified as a change -- and not just a random fluctuation. Let δU_i be the smallest change in U_i that can be so identified. Then, approximately

$$\sum_{i=1}^3 \frac{\delta U_i^2}{U_i} = 1 \quad (3-6)$$

This is, in effect, the condition for unit signal/noise ratio. A change in the stimulus, giving rise to the variation ($\delta U_1, \delta U_2, \delta U_3$), may occur in three independent ways -- a change (ΔL) in luminance, L , a change (Δp) in purity, p , or a change ($\Delta \lambda$) in the dominant wavelength, λ . Let $\delta_L U_i, \delta_p U_i, \delta_\lambda U_i$ ($i=1,2,3$) be the variations in U_i ($i=1,2,3$) for just noticeable differences $\Delta L, \Delta p$ and $\Delta \lambda$ in the stimulus.

From equation (3-5),

$$\begin{aligned} \delta_L U_i &= (\sigma_i + pM_i)\Delta L \\ \delta_p U_i &= LM_i\Delta p \\ \delta_\lambda U_i &= pL \frac{\partial M_i}{\partial \lambda} \Delta \lambda \end{aligned} \quad (3-7)$$

Substituting from (3-7) and (3-5) in (3-6)

$$\begin{aligned} \frac{\Delta L^2}{L} \sum_{i=1}^3 (\sigma_i + pM_i) &= 1 \\ \Delta p^2 L \sum_{i=1}^3 \frac{M_i^2}{\sigma_i + pM_i} &= 1 \\ \Delta \lambda^2 p^2 L \sum_{i=1}^3 \frac{\left(\frac{dM_i}{d\lambda}\right)^2}{\sigma_i + pM_i} &= 1 \end{aligned} \quad (3-8)$$

Let

$$\begin{aligned} L/\Delta L^2 &= F_1(p\lambda) \\ 1/L\Delta p^2 &= F_2(p\lambda) \\ 1/Lp^2\Delta\lambda^2 &= F_3(p\lambda) \end{aligned} \quad (3-9)$$

Then from (3-8) and (3-9),

$$\begin{aligned} F_1(p\lambda) &= \Sigma(\sigma_i + pM_i) \\ F_2(p\lambda) &= \Sigma \frac{M_i^2}{\sigma_i + pM_i} \\ F_3(p\lambda) &= \Sigma \frac{(\frac{dM_i}{d\lambda})^2}{(\sigma_i + pM_i)} \end{aligned} \quad (3-10)$$

Conclusion

The functions of the LHS of equations (3-10) represent the color and brightness discrimination characteristics of the eye. On the RHS occurs the purity coordinate p , the spectral function M_i -- which is, by definition, a function of wavelength only, and constants σ_i . Thus, it is implied in these equations (3-10) that F_1 , F_2 and F_3 (as defined in equations (3-9)) are functions of p and λ only. Thus a first step in verifying these theoretical derivations would be to show that $L/\Delta L^2$, $L\Delta p^2$, $Lp^2\Delta\lambda^2$, -- where ΔL , Δp and $\Delta\lambda$ are the j.n.d's -- are all independent of L .

For example Figure 3-11, taken from Reference 3-7, shows that over a wide range of illumination ΔL is proportional to \sqrt{L} , ie $L/\Delta L^2$ is independent of L .

3.2.6 The Sine Wave Representation of Optical Phenomena

Sine wave analysis is very familiar in electronics. We speak of the sine wave transfer function of networks and the equivalent sine wave representation of signals.

Mathematically, given a function of time $f(t)$, its Fourier Transform is defined as

$$F(\omega) = \int_{-\infty}^{+\infty} f(x) e^{-i\omega t} dt \quad (3-11)$$

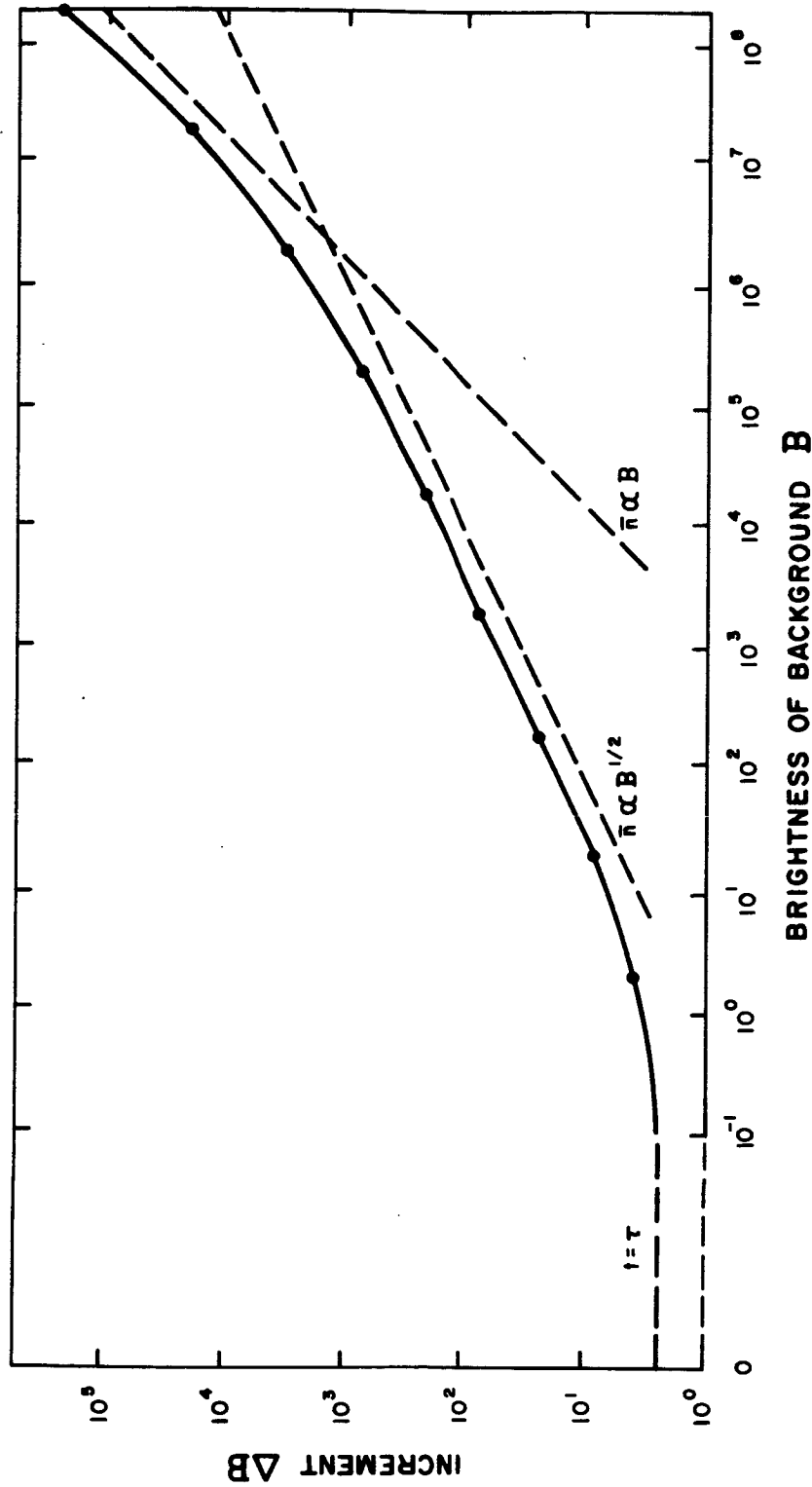


FIG. 3-11
**INCREMENT THRESHOLD AS A
 FUNCTION OF BACKGROUND BRIGHTNESS**
 (REF. 3-7)

The complex derived function $F(\omega)$ completely defines $f(t)$. In fact $f(t)$ can be derived from it by the inverse process

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} d\omega \quad (3-12)$$

Networks that transmit signals can be defined in terms of their response to sine waves of various frequencies. From this transfer function, various pulse responses can be calculated, in particular the impulse response. Let $H(\omega)$ be the sine wave transfer function of a network and let $h(t)$ be its response to a unit impulse.

The Fourier Transform of the unit impulse is

$$\begin{aligned} h(\omega) &= \int_{-\infty}^{+\infty} I(t) e^{-i\omega t} dt, & I(t) &= 0, t \neq 0 \\ &= 1 \int_{-\infty}^{+\infty} I(t) dt = 1, & t &= 0 \end{aligned}$$

Thus the Fourier Transform of the impulse response of the networks is $1/H(\omega)$. Taking the inverse transform of this we have

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(\omega) e^{i\omega t} d\omega \quad (3-13)$$

In the case of an ideal low pass filter of cut-off frequency Ω we can define $H(\omega)$ as follows

$$\begin{aligned} H(\omega) &= 1 & 0 \leq \omega^2 \leq \Omega^2 \\ &= 0 & \omega^2 > \Omega^2 \end{aligned}$$

The phase characteristic (i. e. the argument of $H(\omega)$) is equivalent to a pure time delay of infinite magnitude and can be ignored for practical purposes.

Substituting in equation (3-13)

$$\begin{aligned}
 h(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega t} d\omega \\
 &= \frac{1}{2\pi i t} [e^{i\omega t}]_{-\Omega}^{+\Omega} \\
 &= \frac{1}{\pi t} \sin \Omega t \\
 &= \frac{\Omega}{\pi} \left(\frac{\sin \Omega t}{\Omega t} \right)
 \end{aligned}
 \tag{3-14}$$

This is the familiar $\sin x/x$ function.

Sine wave analysis is used in many other contexts. In some cases it appears a very natural idea, in others as a highly unreal mathematical procedure.

Thus in the case of electromagnetic waves in the visible range there is an intuitive feeling that various spectral components of a light source have a real, separate, physical identity. There is virtually no interest in the actual electric vector waveform of the radiation from a discharge lamp. The light is thought of solely in terms of its spectral content, i.e., in this case it is not $f(t)$ that is of interest but the modulus of $F(\omega)$, $|F(\omega)|$.

In the case of electronics, both frequency domain and time domain concepts are intuitively satisfactory. Sometimes a signal will be observed directly on a CRT, less often its Fourier Transform will be observed on a spectral analyser. Networks are usually defined in terms of their frequency response (e.g., "3db down at 20 kcs"), less often in terms of their pulse response (e.g., 1μ s rise time, 10% overshoot).

An unusual, but very significant application of sine wave techniques in optics describes images and optical equipment. Mathematically a plane image is a two dimensional function of brightness e.g. $b(x,y)$. We may define a two dimensional Fourier Transform of this

$$B(u,v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} b(x,y) e^{-i(ux + vy)} dx dy \tag{3-15}$$

This may be compared with equation (3-11).

In like manner an inverse transform restores the original function

$$b(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} B(u, v) e^{i(xu + vy)} du dv \quad (3-16)$$

The characteristics of optical instruments can also be specified in terms of sine wave analysis. As with networks the "impulse" response may be calculated. This will be the diffraction pattern produced by the instrument of a point source of light. As an example we may consider the Fraunhofer diffraction of a two dimensional aperture:

Let a point source be located (at infinity) on the central axis of an optical system that passes through an aperture A (fig. 3-12). Consider the formation of the image at $+\infty$. Let axes $O_x O_y$ be drawn in the plane of the aperture such that the central axis point is (0,0). Let O_z be the third axis, corresponding with the central axis as shown. Let l, m, n be the direction cosines of any ray directed towards the image plane: (l and m) can be considered as the coordinates of any point in the image. Take an element $dx dy$ at (x, y) in the aperture plane A. The contribution of this element to the image intensity at (l, m) will be a wave

$$e^{i(\omega t + \frac{2\pi}{\lambda} (lx + my))}$$

where w is the electromagnetic wave frequency and $(lx + my)$ is the path length difference between the wave originating from the point (x, y) and the wave coming from the point (0,0) in the aperture plane. This wave function can be written as

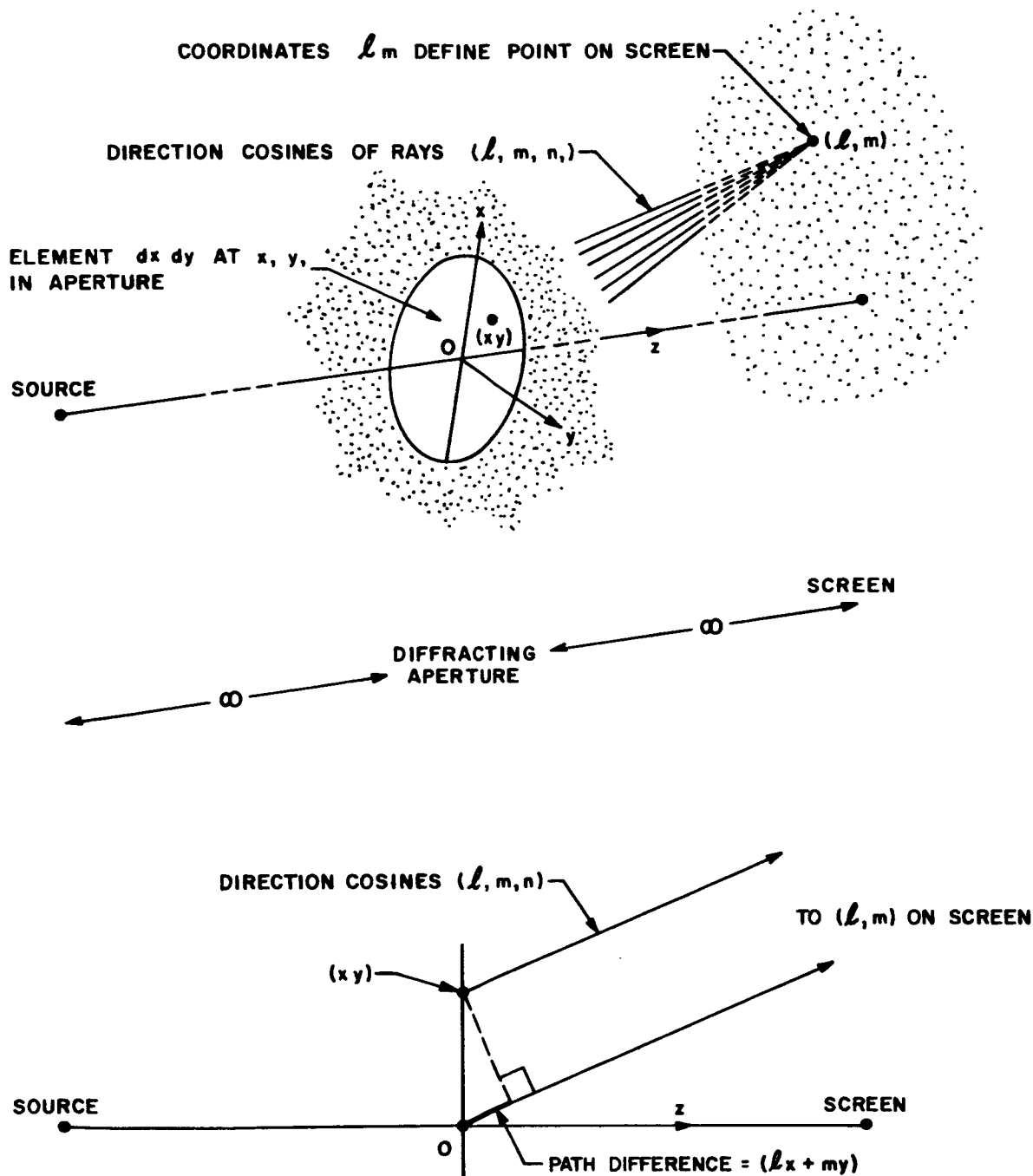
$$e^{i\omega t} e^{i \frac{2\pi}{\lambda} (lx + my)}$$

By superposition, the total intensity of the wave at (l, m) will thus be

$$I(l, m) x e^{i\omega t}$$

where

FIG. 3.12
FORMATION OF DIFFRACTION PATTERN



$$I(l, m) = \iint H(x, y) e^{i \frac{2\pi}{\lambda} (lx + my)} dx dy$$

where $H(x, y)$ defines the transmission character of the diffracting aperture
Putting

$$\frac{2\pi x}{\lambda} = U, \quad \frac{2\pi y}{\lambda} = V$$

This last expression becomes

$$I(l, m) = \frac{\lambda^2}{4\pi^2} \iint e^{i(lu + mv)} H'(u, v) du dv \quad (3-17)$$

where the integral is evaluated over the whole aperture.

In the case of an infinite slit, the problem becomes unidimensional and H takes a simple form

$$\begin{aligned} H(u, v) &= 1 \quad U \leq U^2 \leq \Omega^2 \\ &= 0 \quad U^2 \geq \Omega^2 \end{aligned}$$

$$\text{Where } \Omega = \frac{2\pi L}{\lambda}.$$

Where L is the half width of the slit

Then

$$I(l) = \frac{\lambda}{2\pi} \int_{-\Omega}^{+\Omega} e^{ilu} du$$

This is exactly the same form as in (3-14) and then

$$I(l) = \frac{\lambda \Omega}{\pi} \frac{\text{Sin } \Omega l}{\Omega l}$$

The first zero of this function occurs at $\Omega l = \pi/2$.

$$\text{ie. } \frac{\pi}{\lambda} L l = \frac{\pi}{2}$$

$$\text{ie. } l = \frac{\lambda}{2L}$$

Note that in this analysis I is the amplitude of the electromagnetic wave. The brightness -- as made evident to our eyes and physical instruments -- is proportional to I^2 . Thus if we are to apply the full range of Fourier techniques to optics -- i. e., specify both instruments and images in terms of Fourier Transform -- then the image must be analyzed not in terms of the brightness $b(x,y)$ as in equations 3-15 and 3-16, but in terms of I . However, for the purposes developed in this report the brightness $b(x,y)$ and not the intensity I , is the important physical parameter and for this the sine wave analysis techniques defined by equations 3-15 and 3-16 are appropriate.

3.2.7 A Theory of Peripheral Retinal Sensing

Reference has been made in a recent paper (Ref. 3-8) to the use of Fourier techniques in optics and particularly in physiological optics.

In electronics the method is very well established for linear systems. Thus an electronic signal may be described in the time domain -- e. g. $f(t)$ -- or in the frequency domain, e. g.,

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt$$

Similarly an electric four pole network may be described in terms either of its frequency response (transfer function),

$$G(\omega) = A(\omega) e^{i\phi(\omega)}$$

(where $A(\omega)$ is the amplitude and $\phi(\omega)$ the phase characteristic) or by its impulse response,

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} G(\omega) e^{i\omega t} d\omega$$

Analogously, in optics the "signal" may be regarded as a two dimensional luminance function, $b(x,y)$ (cf $f(t)$). (For the present purpose this function is preferable to the electromagnetic wave amplitude function of which $b(x,y)$ is the square.) The function $b(x,y)$ may be expressed in the two dimensional frequency domain;

$$B(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} b(x, y) e^{-i(xu + yv)} dx dy.$$

$$\text{cf. } \left(F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \right)$$

In some cases the phase component of a signal or network may be unimportant. For example, in audio transmission the phase characteristic of a circuit is generally ignored, and audio signals are usually described in terms only of their amplitude/frequency characteristics. For video transmission, however, both phase and amplitude characteristics are important.

It is suggested that the general distinction between cases in which phase is important and those in which it is not -- a distinction of practical significance in electronics -- may usefully be applied to explain, or provide a basis for an explanation of, the difference between foveal and peripheral vision.

Some such explanation would appear to be necessary since peripheral vision is not simply a less acute form of foveal vision but is of a somewhat different nature. The peripheral retina is sensitive to relatively high frequency spatial detail yet has a very poor ability to perceive form -- for example alphabetic characters. In foveal vision, on the other hand, the sensitivity to high frequency detail and letter sensing acuity appear to be well correlated.

Although the sensory capability of the fovea is much greater than that of any other part of the retina of the same angular size, the total sensory information capacity of the peripheral retina must be much greater than that of the fovea. This is evidenced, for example, by the fact that most of the optic nerve fibers terminate in the peripheral retina. Thus, from the point of view of information channel capacity, peripheral vision would appear to be much more important than foveal vision. It is logical to seek meaningful descriptions of the sensory characteristics of this very important part of the visual sense.

In the case of foveal vision, these characteristics (neglecting color) are essentially acuity and liminal brightness discrimination. An external image having a 'grain' and contrast corresponding to these foveal parameters

will generally be judged as being of good quality and all the detail that is to be seen in the image can be seen. There is, in other words, an efficient match between the information content of the image and the information sensing characteristics of the fovea. A similar match cannot be made with the peripheral retina. When the grain size in the image is adjusted to match the peripheral grating acuity -- ie. sensitivity to sharp edges -- there is far more detail in the image than can be seen by the peripheral retina. For example, alphabetic characters five times the grain size will be clearly defined in the image yet cannot be identified by the peripheral retina. For example, alphabetic characters five times the grain size will be clearly defined in the image yet cannot be identified by the peripheral retina. If, on the other hand, the grain size is adjusted so that it does correspond with the smallest test letter than can be identified, then the image will appear fuzzy, due to the loss of sharp edges. In other words, it is not possible to match an image efficiently to the peripheral retina using the same parameters that were adequate to form a match with the foveal retina.

A simple mathematical model is proposed to explain the difference between peripheral and foveal vision.

An analogy may be drawn between peripheral vision and the aural sense. The input stimulus to the ear is a one dimensional pressure time function $p(t)$, and to the eye a two dimensional luminance function $b(x,y)$. The ear can detect high frequencies, e.g. 10 kc

$$p(t) = \sin \omega t, \omega = 2\pi 10^4 \text{ cps}$$

ie., it has a time resolution of 100, μ s. Similarly the peripheral retina can resolve relatively fine gratings e.g., 10 minutes of arc 20° from the fovea,

$$b(x,y) = \sin \omega x \quad \omega = 12\pi \text{ lines per degrees}$$

ie., it has a spatial resolution of 10 minutes of arc. However, the ear can detect the 'form' of the function $p(t)$ -- well enough, for example, to enable Morse Code signals to be recognized -- only up to a time resolution of the order of 100 ms. Similarly the peripheral retina, at 20° from the fovea, can detect the form of $b(x,y)$ -- well enough to enable alphabetic characters to be recognized -- only up to a spatial resolution of the order of several degrees.

It is well known that the ear does not respond to the function $p(t)$ itself, but rather to its power spectral content, averaged over periods of time of the order of 100 ms -- ie. to a function of the form,

$$P(\omega, T) = \left| \int_{T - \tau}^{T + \tau} p(t) e^{i\omega t} dt \right|$$

where τ is the order of 100 ms. It is hypothesized that, analogously, the peripheral retina responds not to $b(x, y)$ -- as does the fovea -- but to the spatial power spectral content of $b(x, y)$ averaged over areas of the order of a few square degrees, ie. to a function of the form,

$$B(u, v, X, Y) = \left| \iint_{R(X, Y)} b(x, y) e^{iux} e^{ivy} dx dy \right|$$

where the double integral extends over a closed area R , the position of which is defined by X, Y . In the analogy with the ear, R is the analog of τ and X, Y the analog of T . τ is totally or largely independent of T but R is probably a strong function of X, Y .

The advantage of the theory is that it offers a simple mathematical model to explain, in qualitative terms, the difference between peripheral and foveal vision. It is hypothesized that the peripheral retina operates analogously to the ear in that it senses the power spectral content of the stimulus -- not the stimulus itself. At best, this may prove to be only a crude approximation to reality, but it may nevertheless be of value as a framework on which to build a more accurate model using experimental quantitative data.

The spectral analysis, hypothesized for peripheral sensing could actually be autocorrelation of the function $b(x, y)$ by the retinal neural networks connecting the 10^8 photoreceptors to the 10^6 optic nerve fibers. Alternatively instead of autocorrelation of $b(x, y)$ with itself, the process might be cross correlation by $b(x, y, t)$ with $b(x, y, t + \delta t)$. This would introduce the time variable, indicating a possible connection between the acute peripheral sense of motion and movement, and the hypothesized spatial spectral analysis function of the retina.

3.2.8 Signal Processing Techniques of Nature

In certain cases the human senses do not absorb all the information contained in the 'signal' impinging on them. For example, the visual sense responds to electromagnetic radiation over a certain band. However, no information is obtained about the actual wave intensity as a function of time. Thus, if an electromagnetic wave defined by the electric vector

$$\underline{E} = \underline{E}(t)$$

is incident at the eye the only information that is absorbed is a limited amount of spectral information, averaged over periods of time that are very long compared to the period of vibration of the electromagnetic wave. For all practical purposes the actual stimulus can be regarded as a power spectral distribution, defined by

$$P(\omega) = \left| \int_{-\infty}^{+\infty} E(t) e^{-i\omega t} dt \right|$$

It is usual to consider this function in terms of wavelength λ , where,

$$\omega = \frac{2\pi c}{\lambda}$$

It is important to note, however, that only a very limited amount of information about $P(\omega)$ is derived. The results of color matching experiments indicate that $P(\omega)$ is 'sampled' at three different frequency bands and the only information obtained is the value of these three samples. It is for this reason that stimuli having completely different spectral distributions may appear, subjectively, to be identical.

Another example can probably be seen in the case of peripheral vision, where -- as discussed elsewhere -- there is evidence to suggest that the subjective response is to the spatial spectral power density function averaged over areas of the order of a few square degrees -- depending on the distance from the fovea.

Thirdly, it is well known that the ear does not respond to the pressure/time function at the ear but rather to its power spectral content, averaged over periods of the order of 1/10 sec.

These three important examples of spectral sensing in nature suggest the possible existence of a general principle -- worthy of bionic study -- that might be applied in engineering science.

Let it be supposed that a sensing process involves detection of the function $f(x)$. From the point of view of pure mathematics it is perfectly legitimate to consider the function defined in this simple way. Practically, in

any case involving physical observation, there will be a finite degree of accuracy -- ie in any attempt to measure $f(x)$ the result will be

$$f(x) + n(x)$$

where $n(x)$ can be described as a noise function. Furthermore, measurements cannot be made at infinitesimally small sampling intervals. Thus the result of any sensing process may be described as a set of numbers

$$\sigma_r = f(x_r) + n(x_r)$$

where $x_{r+1} - x_r$ is the sampling interval and r is an index specifying a particular sampling interval.

It may be shown that any function $\phi(x)$, that contains no frequency components above Ω ; that is, such that

$$\int_{-\infty}^{+\infty} \phi(x) e^{-i\omega x} dx = 0 \quad \omega > \Omega$$

is completely defined by the "time series" η_r where

$$\eta_r = \phi(x_r)$$

and

$$x_{r+1} - x_r = \pi / \Omega$$

It may further be shown that, under the stated conditions,

$$\phi(x) = \sum_{r=-\infty}^{+\infty} \eta_r \frac{\sin \pi \left(\frac{x}{\lambda} - r \right)}{\pi \left(\frac{x}{\lambda} - r \right)} \quad (3-18)$$

where

$$\lambda = \pi / \Omega$$

It does not follow that the samples σ_r of an unrestricted function $f(x)$ define a band limited version of this function, according to equation 1. However, a function of the form of equation 3-18 represents a good approximation to the band limited version of the function $f(x)$, and probably represents the best possible approximation to the function $F(x)$.

The Fourier Transform of the function derived from the samples σ_r of $f(x)$ can be shown to be

$$F(\omega) = \sum_{-\infty}^{+\infty} \sigma_r e^{-ir\omega\tau} \quad \left(0 \leq \omega \leq \pi/\tau\right)$$

The power spectral density is given by

$$\begin{aligned} P(\omega) &= \left| F(\omega) \right|^2 = F(\omega) \overline{F(\omega)} \\ &= \sum_{-\infty}^{+\infty} \sigma_r e^{-ir\omega\tau} \sum_{-\infty}^{+\infty} \sigma_r e^{+ir\omega\tau} \\ &= \sum_0^{\infty} U_p \cos p\omega\tau \end{aligned}$$

$$\text{where } U_p = \begin{cases} 2 \sum_{-\infty}^{+\infty} \sigma_r \sigma_{r+p} & p \neq 0 \\ \sum_{-\infty}^{+\infty} \sigma_r^2 & p = 0 \end{cases}$$

Let us consider the case now of a function $f(x)$ defined only over a finite range of x ; corresponding to just $N+1$ samples σ_r .

$$\text{Then } f(x) = \sum_{r=0}^N \sigma_r \frac{\sin \pi \left(\frac{x}{\lambda} - r \right)}{\pi \left(\frac{x}{\lambda} - r \right)}$$

$$\text{and } P(\omega) = \sum_0^N U_p \cos p\omega\tau \quad (3-19)$$

$$\text{where } U_p = 2 \sum_{r=0}^N \sigma_r \sigma_{r+p} \quad (3-20)$$

Let us suppose that some sensing mechanism can give complete information concerning $P(\omega)$. Then it follows from equation 3-19 that all the coefficients U_p $p=0,1,2,\dots,N$ are determined by the sensing process. Then the $N+1$ equations (3-20) can be solved uniquely for the σ_r except for an ambiguity of sign

and time. That is, if the set σ_r is a solution then $-\sigma_r$ and $\pm\sigma_{N-r}$ are also solutions.

The important conclusion to be drawn is that it takes just as many parameters to define $P(\omega)$ as $f(x)$, and correspondingly -- except for the two ambiguities in sign and time -- $f(x)$ is completely defined when $P(\omega)$ is completely defined. This is in spite of the fact that $P(\omega)$ -- being the power spectral function -- contains no phase information.

It is generally assumed that the ear senses the spectral composition of sound energy. If the spectral density were fully and accurately sensed it can be seen from the above analysis that almost all the information about the actual pressure time function could be derived. Since it is well known that this is not the case -- and that it is very unlikely that the brain fails to extract as much information as possible from the sensor signals -- it follows that, in the case of the aural sense, there must be incomplete spectral sensing.

This is certainly the case for color vision. The spectral distribution of luminous energy is not at all accurately sensed by the color mechanism. The luminous energy spectral distribution $P(\omega)$ may be said to be "sampled" at three (broad) values of ω . This fact, in turn, manifests itself in the color matching phenomena. If sound sensing is in fact also incomplete spectral sensing why is it that similar sound matching phenomena are not observed? That is, it should be possible to match any sound at all (of a given duration) with suitable combinations of a finite number of pure tones (of equal duration).

It is postulated that the reason that no such phenomena have been observed is that the spectral analysis performed by the aural sense is not fixed, as in the case of color vision, but is variable. That is, although only a finite, restricted, amount of information can be obtained, at any one time, about the sound spectrum $P(\omega)$ -- for example the energy falling in certain bands -- the kind of information that is selected can be modified -- eg by shifting the sampling bands.

Conclusion

Examples have been given of the use in nature of incomplete spectral sensing of a function. The fact that sufficient information can be extracted in this way to enable the organism to function effectively in its environment

suggests that similar sensing methods might usefully be employed in engineering.

3.3 Engineering Applications

3.3.1 Introduction

This section is devoted to a discussion in some detail of the engineering applications suggested in the previous report (see Ref. 3-1).

The engineering applications derived from the current study are not as readily defined in detail at this time, since they relate to a much more difficult part of the visual process, and also the systems to which any such applications might be tied are not fully developed. Essentially these systems are of the kind that would operate, automatically, with visual patterns -- just as simple photocell systems now operate with the "one-bit" average brightness of a scene over one, defined, region. The development of these pattern sensing optical systems may prove to be based on the fact that the visual sense does not -- except for the very small foveal region -- sense the point by point brightness function of the viewed scene but, possibly, a limited amount of the spatial spectral content of this function.

3.3.2 The Applications of Oculometry

An engineering analysis of the human eye system previously made at Allied Research (see Ref. 3-1) showed that the oculomotor apparatus is of fundamental significance to the total operation of the visual sense. Correspondingly, a number of important applications in (visual) human engineering were recognized for oculometry -- the instrumentation technique to measure the direction of the axis of regard of the eye. In general these applications result in a substantial improvement in man's visual capacity to act as an element in a total man/machine system.

The very non-uniform nature of the eye's perceptive field requires that it function in some respects like a PPI radar antenna. (This is a narrow beam antenna that is continuously rotated, at about one revolution per second, usually around a vertical axis.) In order to see detail clearly, the eye must be mechanically oriented so that the image of this detail falls over the very small part of the retina -- the fovea -- which has maximum perceptive capability. The angular accuracy to which it must, and does, track viewed detail is of the order of 5 - 10 minutes of arc -- which is close to maximum foveal acuity (approximately one minute of arc). Spatial information is derived by the brain from the oculomotor system of the eye in much the same way that

the PPI radar picture information is derived through sensing of the radar antenna's angular position. The radar beam, in itself, is not capable of yielding complete spatial detail -- correspondingly the retina is not capable of accurate sensing of large-scale detail. Not only is oculomotor tracking very accurate, it is also extremely rapid -- in fact the neuromuscular system responsible for the control stabilization and motion of the eye must be rated as one of the most sophisticated of such systems in the body.

Although the foveal part of the retina is used to perceive detail, the peripheral retina is of great importance in the total visual sense. Its perceptive capability per unit angular area is quite poor - compared to that of the fovea - but the total amount of information absorbed by the peripheral retina at any time is much greater than that absorbed by the fovea. This is clear both from the known structure of the retina (e.g., the distribution of optic nerve fibers over the retina) and also behaviourly, in terms of the loss of visual capability when the peripheral retina is blind.

A principal application of oculometry is in certain high performance control systems, in which only a human operator can perform the control task. For example, the guidance of an air-to-ground missile against targets that have a characteristic signature only the human eye/brain system can detect. Oculometry can be applied to increase the performance of the human operator from the limit now set by his ability for conscious action to the much higher limit set only by his ability for visual perception. Such improved performance will be of significant value in high performance systems that must depend on human control action.

Another application of oculometry is a television system in which picture quality is non-uniform, and matched to the perceptive characteristics of the retina. In this way a remote visual capability approaching that of normal vision with the naked eye may be attained with bandwidth economy.

Several oculometric methods have been used in the last few years in research projects simply to measure eye motion. As indicated above, however, a number of applications exist for an oculometer, not simply as an open loop measuring device, but rather as part of a closed loop system. Development of these applications requires that the technique of oculometry be advanced from its present experimental laboratory status to operational form. Additionally, research is required into the operation of systems employing the human oculomotor system in an external closed loop.

The two principle applications are discussed in the remainder of this section. The instrument itself - the oculometer - is discussed in Section 3.3.3.

A. Command Control

There are many control tasks assigned to a human operator, usually because of his unique pattern sensing capability. Thus, for example, a human is generally required to control a missile tracking telescope (or camera). When the missile tends to move out of the field of view, the operator senses this, and by conscious neuromuscular action changes the orientation of the camera in order to bring the missile back into the center of the field of view. All such hand-eye coordination control tasks contain a closed feedback loop to which the man contributes the steps:

- a. perception of the visual situation
- b. neuro-muscular control action - usually some form of manual control.

Closing the loop is the inanimate part of the system, which translates the neuro-muscular control action into changes in the visual situation (See Fig. 3-13).

1. Most of the sluggishness and inaccuracy of the operator is contained in step (b) above, i. e., in his conscious neuro-muscular control action.
2. Spatial information can be read out directly from a human by means of oculometry. In this way the most inefficient and undesirable part of the operator's action is eliminated.

The benefit of this application of oculometry is that when, as is often the case, overall system performance is limited by the human operator characteristics, this performance can be improved by oculometric readout from the human operator.

As in any feedback system, sluggishness and inaccuracy in the feedback loop will result in deviation of the controlled quantity from its desired value. In many human operator systems, the sluggishness and inaccuracy of the human is the limiting feature of the system.

Figure 3-14 illustrates how oculometry can be applied to the problem of controlling a man-guided missile. Figure 3-15 shows how the same system must operate without oculometric control. It can be seen that with oculometric control the feedback loop is much shorter.

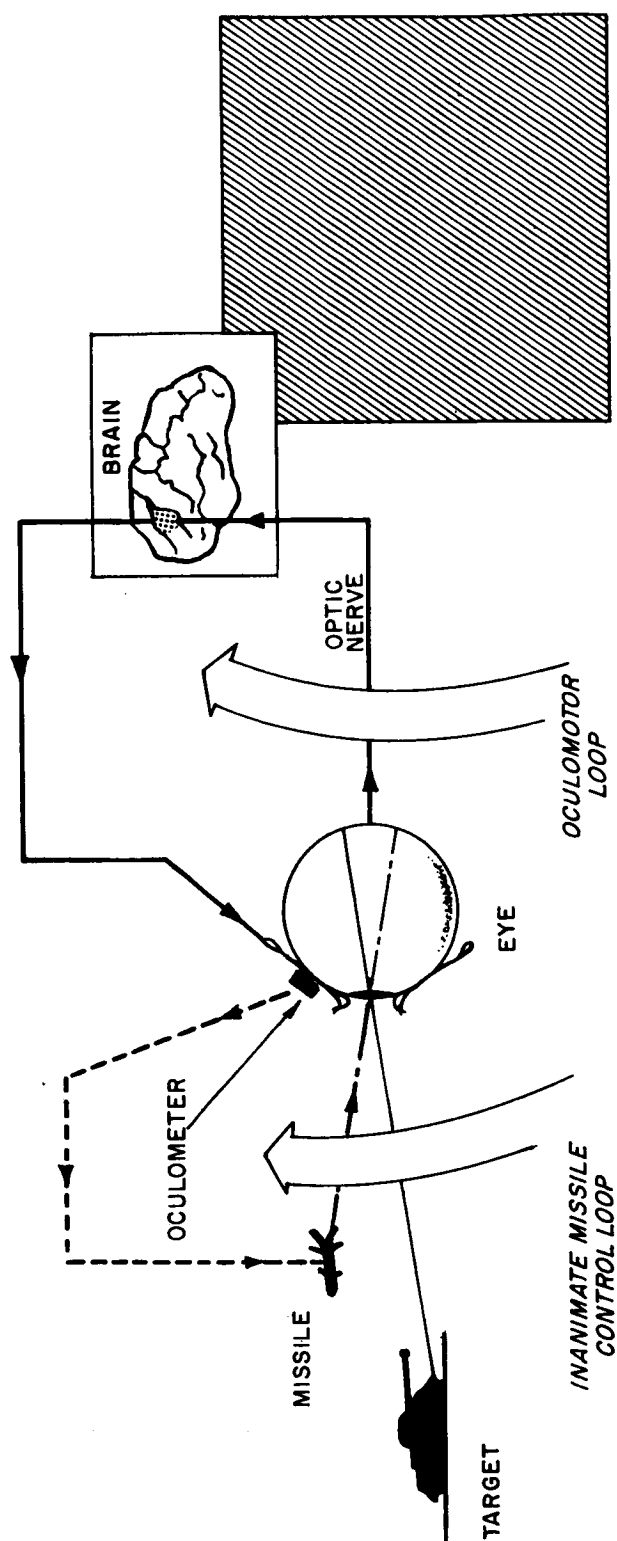


FIG. 3-14
COMMAND CONTROL SYSTEM WITH OCULOMETER

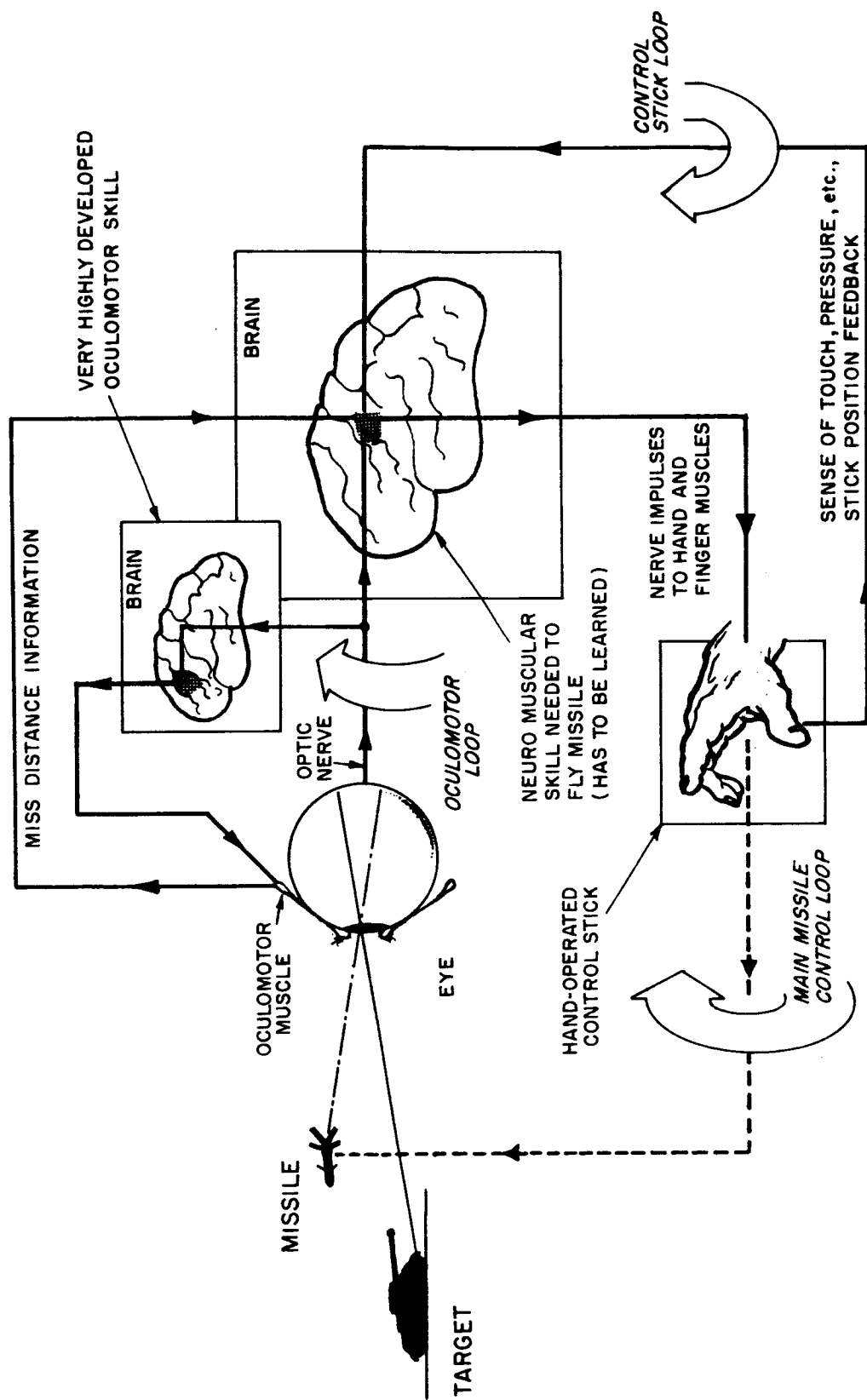


FIG. 3-15

TYPICAL COMMAND CONTROL SYSTEM

There are other possible applications in a wide range of human control tasks, e.g., orbital docking, readout from a radar PPI display, weapon fire control, etc.

The ability of the eye to function directly as a control element as described above depends on:

1. the static pointing accuracy of the eye
2. the dynamic performance of the oculomotor servo system.

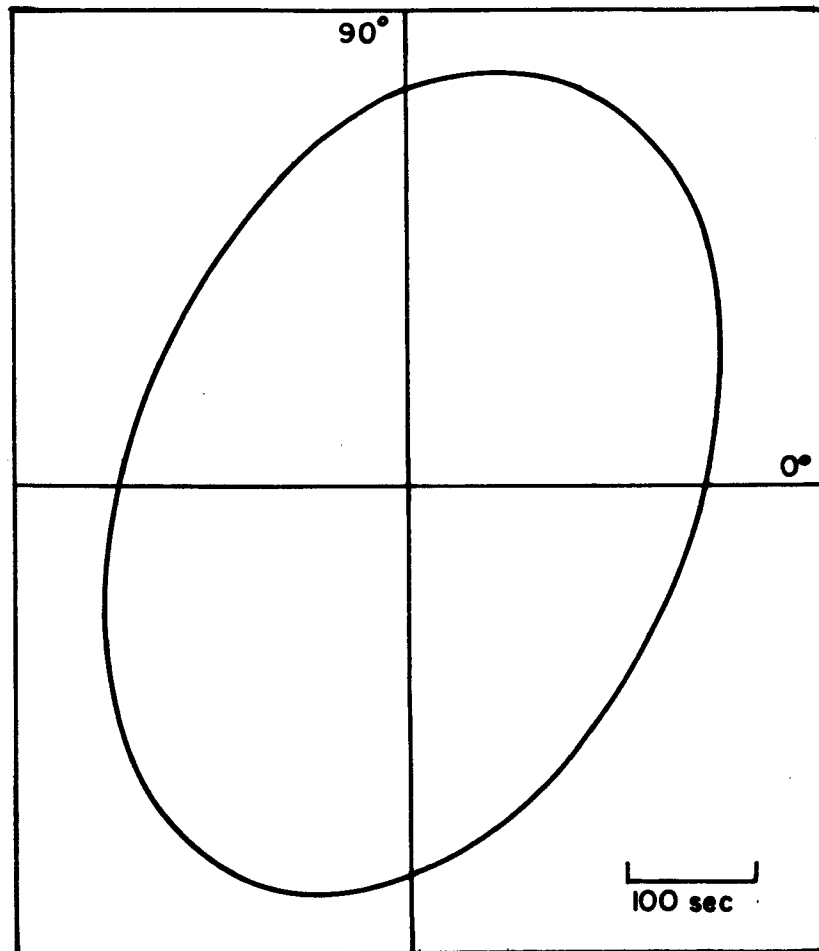
Under conditions of steady fixation the nominal axis of regard is directed to an accuracy of the order of 5 - 10 minutes of arc. This is illustrated in Figure 3-16, taken from Reference 3-9, which shows the one region of the position of the axis of regard during a 30 second period of fixation. Figure 3-17 taken from Reference 3-10 shows that the median amplitude of eye movement after a one second period of fixation is 3 minutes of arc.

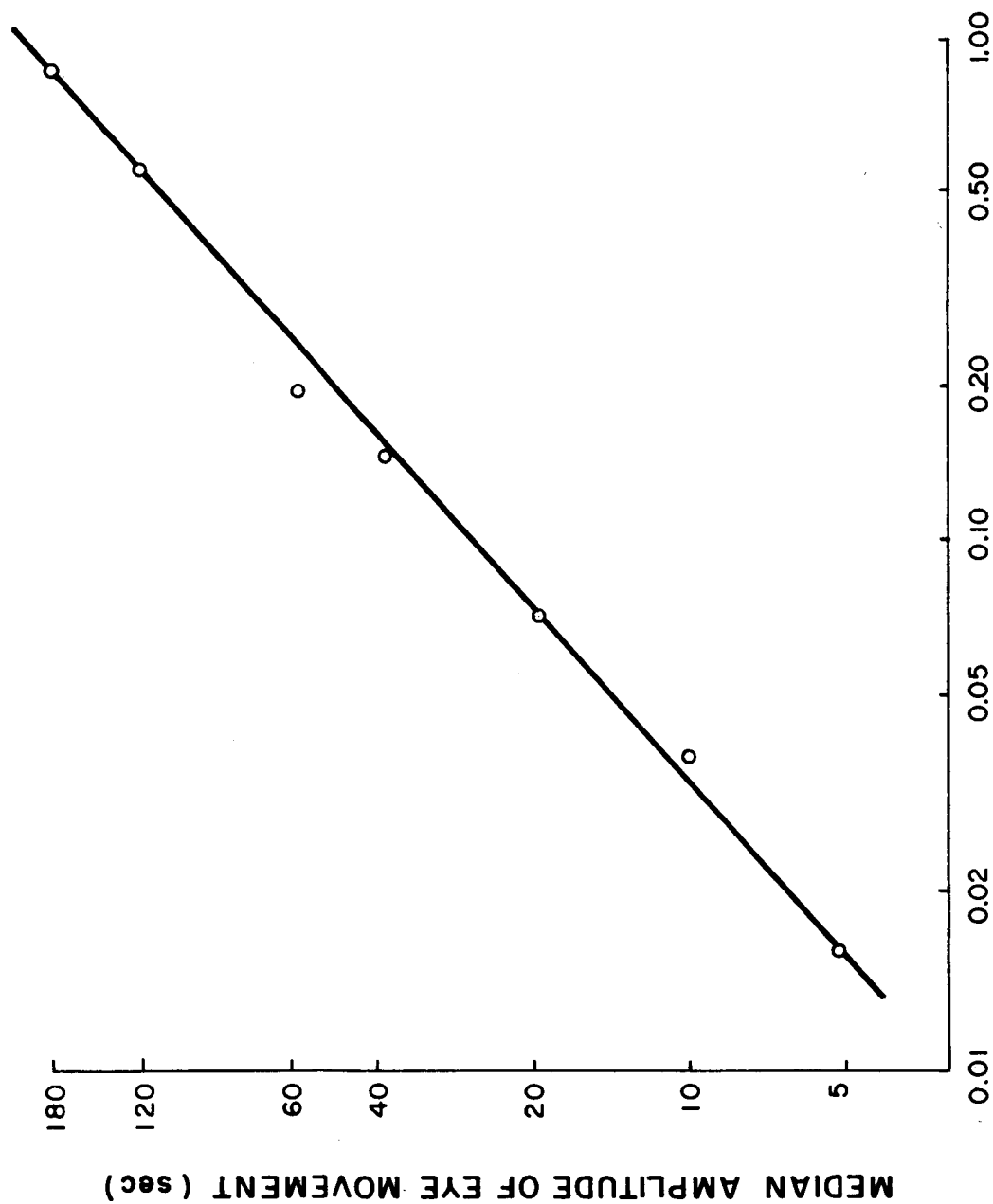
The dynamic characteristics of the eye have been measured. For a 30° movement Westheimer found that the eyes could track, reasonably accurately, a 3 cps sine motion (Ref. 3-11). For constant velocity targets he reports that, after a delay of 150 - 250 ms, the eyes move with the velocity of the target. Small saccadic motions of the order of 1° then occur back and forth around the target. The transfer function of the physiological servo system controlling eye motion has been measured (Ref. 3-12). The results are shown in Figure 3-18. Gain and phase characteristics are given for both predictable and unpredictable target motions (of 5° amplitude). It has been shown that the oculometer apparatus is equivalent to a sampled data control system (Ref. 3-13).

There is a substantial body of data in the literature indicating that the eye is controlled in steady fixation to a high degree of accuracy (5 - 10 minutes of arc -- i.e., about 0.2 percent of the total dynamic range of possible eye motion) and also that its dynamic characteristics are comparable, for example, to those of IR and radar homing heads. Thus the human eye/brain combination can be thought of as a target seeking device with a performance similar to simple automatic homing devices but with the unique feature that it can operate with a complex signature -- such as an "enemy" tank. Automatic devices can operate only with simple signatures

FIG.3-16

TYPICAL FIXATION AREA DURING
A 30 SEC PERIOD OF TIME

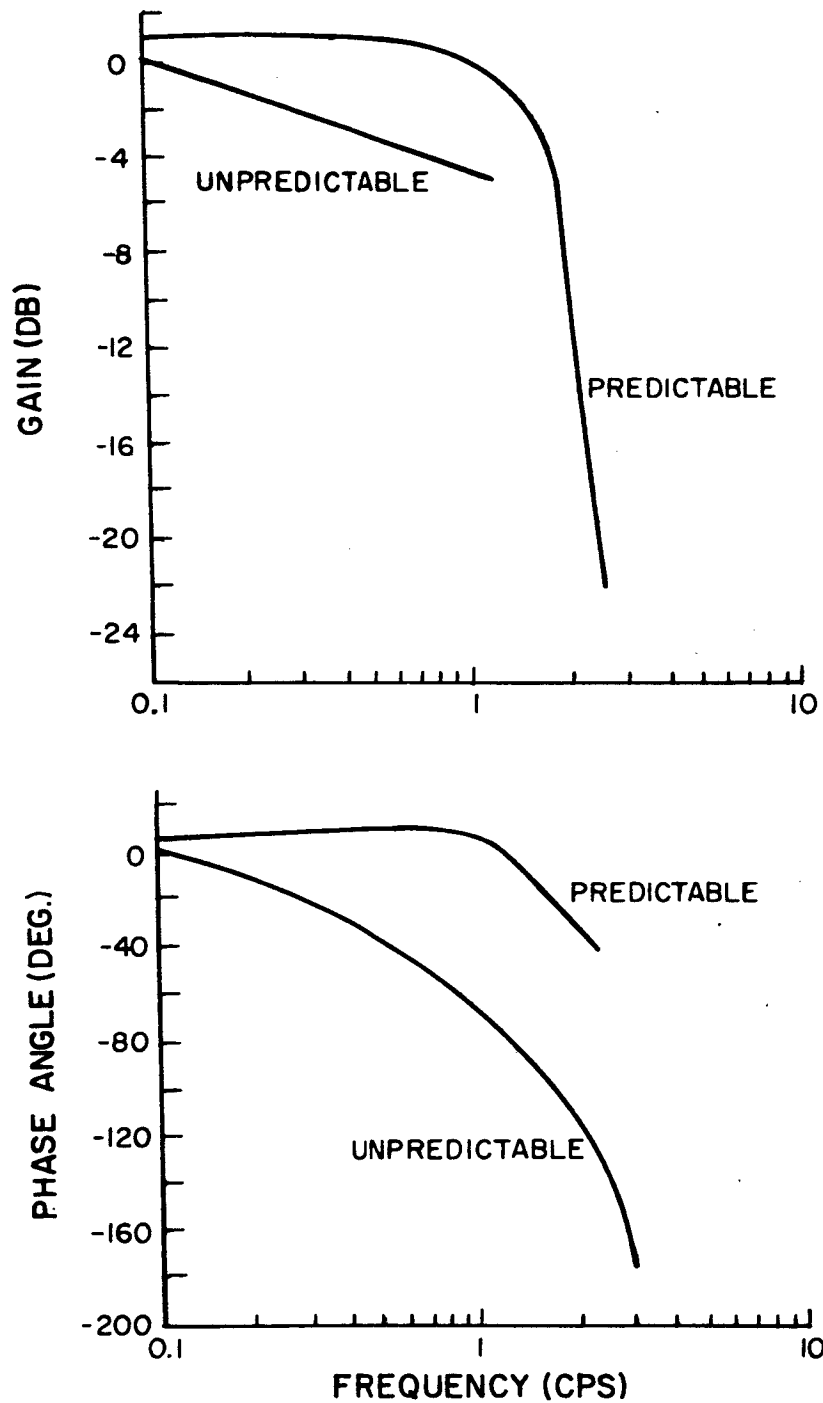




EYE MOTION DURING STEADY FIXATION
FIG. 3-17
(TAKEN FROM REF. 3-10)

FIG. 3-18

GAIN AND PHASE RELATIONSHIPS FOR CONTINUOUS PREDICTABLE AND UNPREDICTABLE TARGET MOTIONS



such as infrared radiation. (Automatic pattern sensing devices are being developed but are presently at an elementary stage.) In the normal use of the human operator (i. e., without oculometric read out from the eye) the same unique pattern sensing function exists but the dynamic characteristics are substantially inferior to the automatic homing devices.

System Application - TV Guidance

A typical control application of oculometry is to meet a requirement for an air-to-surface missile capable of achieving accuracies of the order of several feet against point targets such as armored vehicles. It is desired that such a weapon be fired from a high speed fighter at ranges of several thousand yards and that evasive maneuvers of the fighter be initiated immediately after missile firing, to avoid the tactic by the attacking aircraft of flying the line of sight to the target behind the missile.

Of prime importance, in meeting the above requirement, is the development of a suitable guidance system capable of discerning and tracking the target. A very fundamental question in relation to this problem is whether it will be possible to sense the target automatically, (say from the missile) or whether the guidance system must include a human operator. Heretofore it has not, in general, proved feasible to guide battlefield missiles automatically onto point targets. Future systems, such as pattern recognition and automatic lockon devices for homing in on the target, may be some time coming. Present systems which utilize the human operator are objectionable because of sluggishness and inaccuracies caused by conscious neuro-muscular control.

As a result of this study, a new approach to human operator utilization has been formulated which avoids the sluggishness and inaccurate neuro-muscular control action inherent in present hand (stick) controlled systems. Broadly, the concept requires the operator simply to look at the target: while he does this the direction of the axis of regard of his eye is monitored. This monitored data is converted into continuous real time target bearing signals that are transmitted to the missile as it approaches the target.

The technique of measurement of eyeball orientation is referred to here as oculometry, and the instrument to measure such orientation as an oculometer.

Suppose that a practical oculometer exists, capable of measuring eye-ball orientation under operational conditions to within five minutes of arc, over a range of eye motion of $\pm 10^\circ$ vertically and horizontally from a reference axis.

The following is presented as an illustrative example of how such an oculometer could be used in an air-to-ground guided missile system. The advantages to be gained by using oculometry in this particular example are pointed out: they are typical of the benefits resulting from the application of oculometry to many human operator systems.

The pilot (P), flying the missile carrying plane, observes a target (T) at any range between, say, half a mile to several miles (Figure 3-19). The pilot wears a special helmet (H) fitted with a TV viewing-device (Figure 3-20). He can view normally through a half silvered mirror (M) which may also project an image of the TV screen (V) into his eye.

The plane carries a missile (M) which is fitted with a TV camera (C) and a steerable optical system G (Figure 3-21). This latter consists of a gimballled mirror/lens system capable of being oriented on command from the oculometer in the pilot's helmet. In this way the TV central axis (X) can be pointed anywhere within, say, 20° of the missile central axis. A radio communication link (L) exists between the missile and mother plane. Video signals are transmitted from the missile, and optical axis steering commands are sent from the plane to the missile.

Before the actual attack is initiated, the pilot will have normal vision to view his instruments and to look out of the cockpit window. During this time the attitude of the helmet, relative to the plane, will be continuously monitored. This information will be used to control the direction of the optical axis of the camera in the missile that is being carried under the wing of the aircraft. As soon as the pilot spots a target (Figure 3-22) that he wishes to attack he will press the fire button, keep his eye on the target, and - if he has not already done so - move his head so that the circular reticule, corresponding to the edge of the TV field of view, contains the target. As soon as the fire button is pressed the pilot's eye direction, relative to the center of the TV field, will be monitored by the oculometer. As the pilot moves his head to make the reticule contain the target, the angle detected by the oculometer will decrease until it is less than the half width of the TV system, i. e., about 5° . As soon

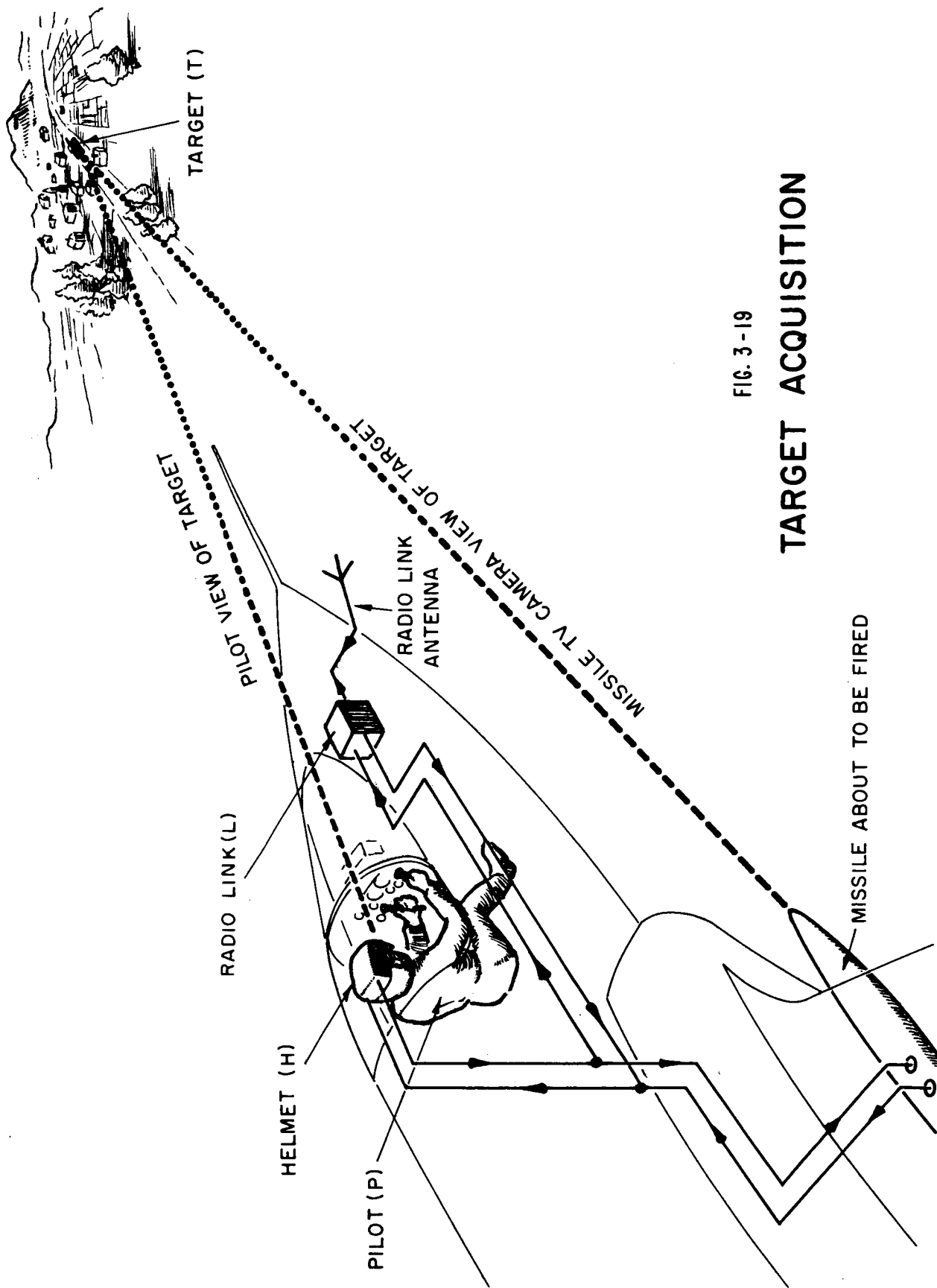
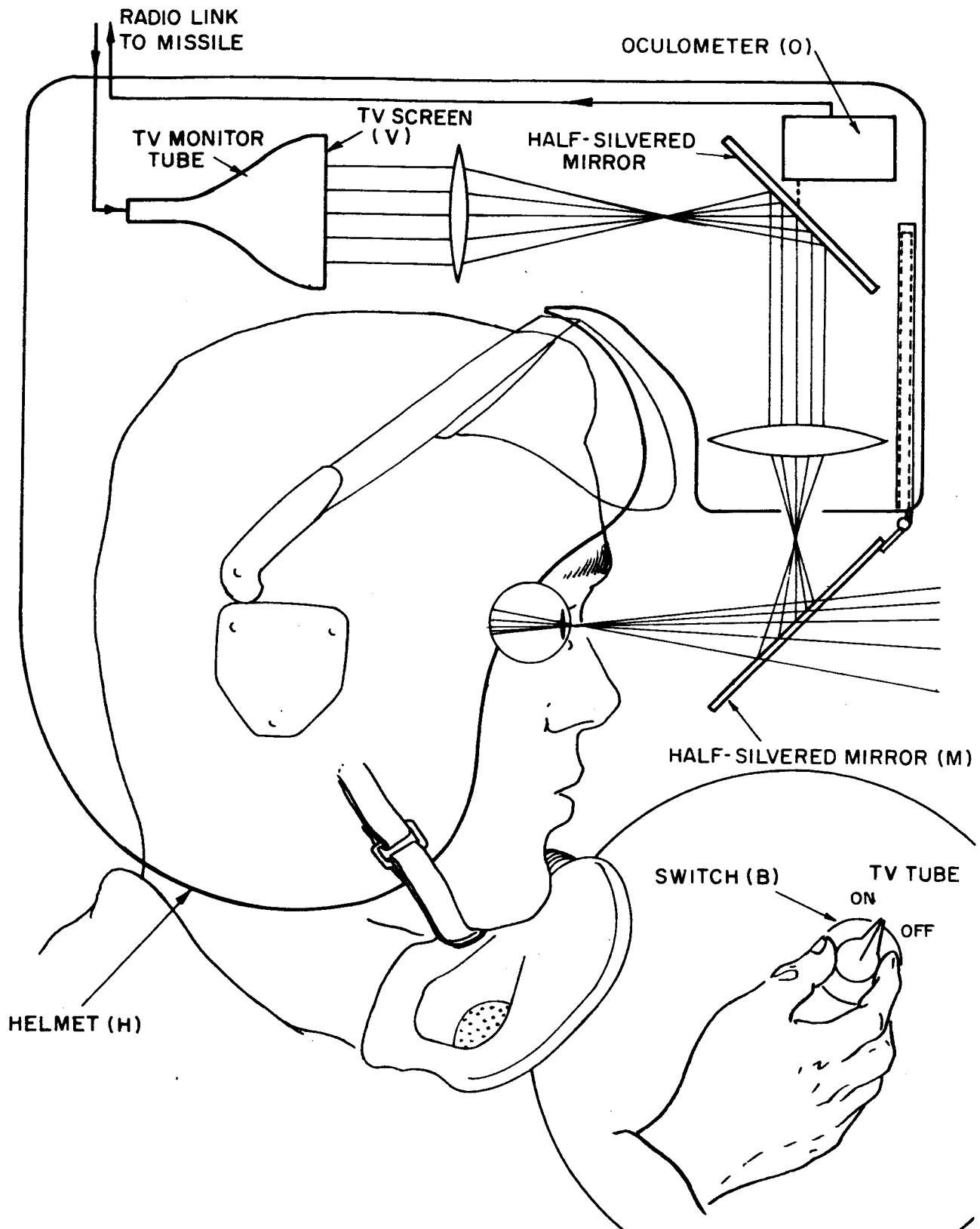


FIG. 3-19

TARGET ACQUISITION

FIG. 3-20
PILOT'S HELMET



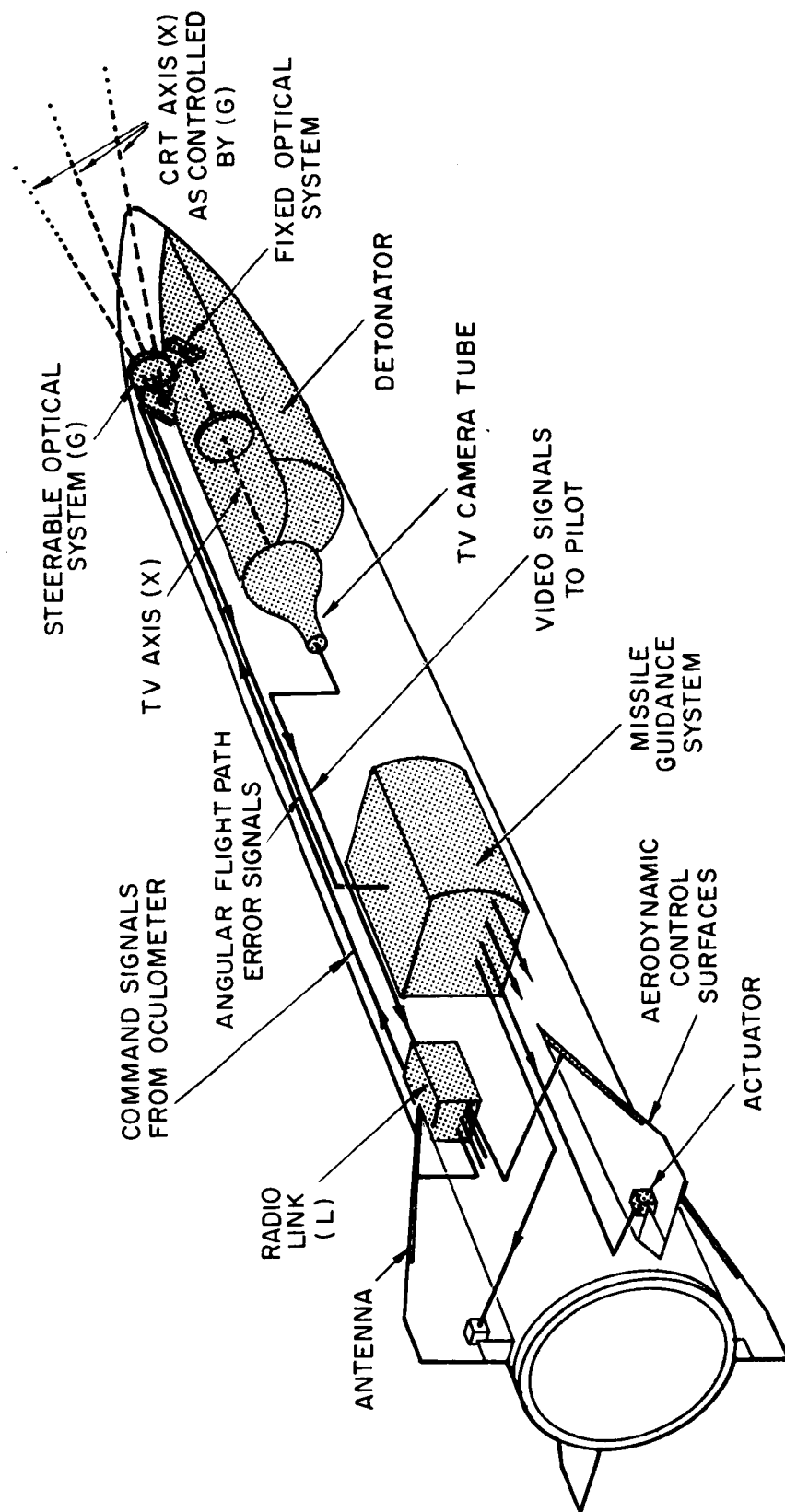


FIG. 3-21
TV GUIDED MISSILE

as it falls to this value the following three things will automatically happen (Figure 3-23):

1. The missile will be fired.
2. The television picture coming from the missile will be immediately displayed to the pilot by means of the helmet mounted TV screen. The brightness of this screen will be sufficient to wash out the pilot's direct view -- by transmission through the half silvered mirror -- over an area equal to the angular size of the TV screen, i. e., about 10° . Over the rest of his visual field the pilot will, as before, be able to see through the half silvered mirror the visual scene lying in front of his eyes.
3. The camera axis in the missile -- which up to this moment will have been slaved to the angular attitude of the helmet -- will be controlled by signals coming from the oculometer measuring the direction of pointing of the pilot's eye. This control loop will be designed to bring whatever detail the pilot chooses to look at (i. e., the target) to the center of the TV screen. When this control action has been accomplished, the TV camera axis in the missile will necessarily be pointing at the target. Thus the angle between the camera axis and the missile will be related to the current flight path error of the missile, and may be used to activate a missile-borne homing guidance system to bring the missile into a collision course with the target.

A very important aspect of this system is that, at the instant the missile is fired, the detail contained in the TV picture will correspond almost exactly with that part of the external scene washed out by the TV image (Figure 3-23). There will, therefore, be no interruption of the pilot's view of the target. At the instant before the missile is fired he will see it by direct view, and immediately after firing -- in almost exactly the same place -- by TV. That is, it will not be necessary for him to change the direction of pointing of his eye in order to see the target in the TV screen. There will be no acquisition phase of the guidance function subsequent to his pressing the firing button.

As soon as the missile has been fired the pilot's only guidance task will be to keep his eye on the target -- i. e., to "look-at" the target -- as it appears in his headset TV screen (Figure 3-24).

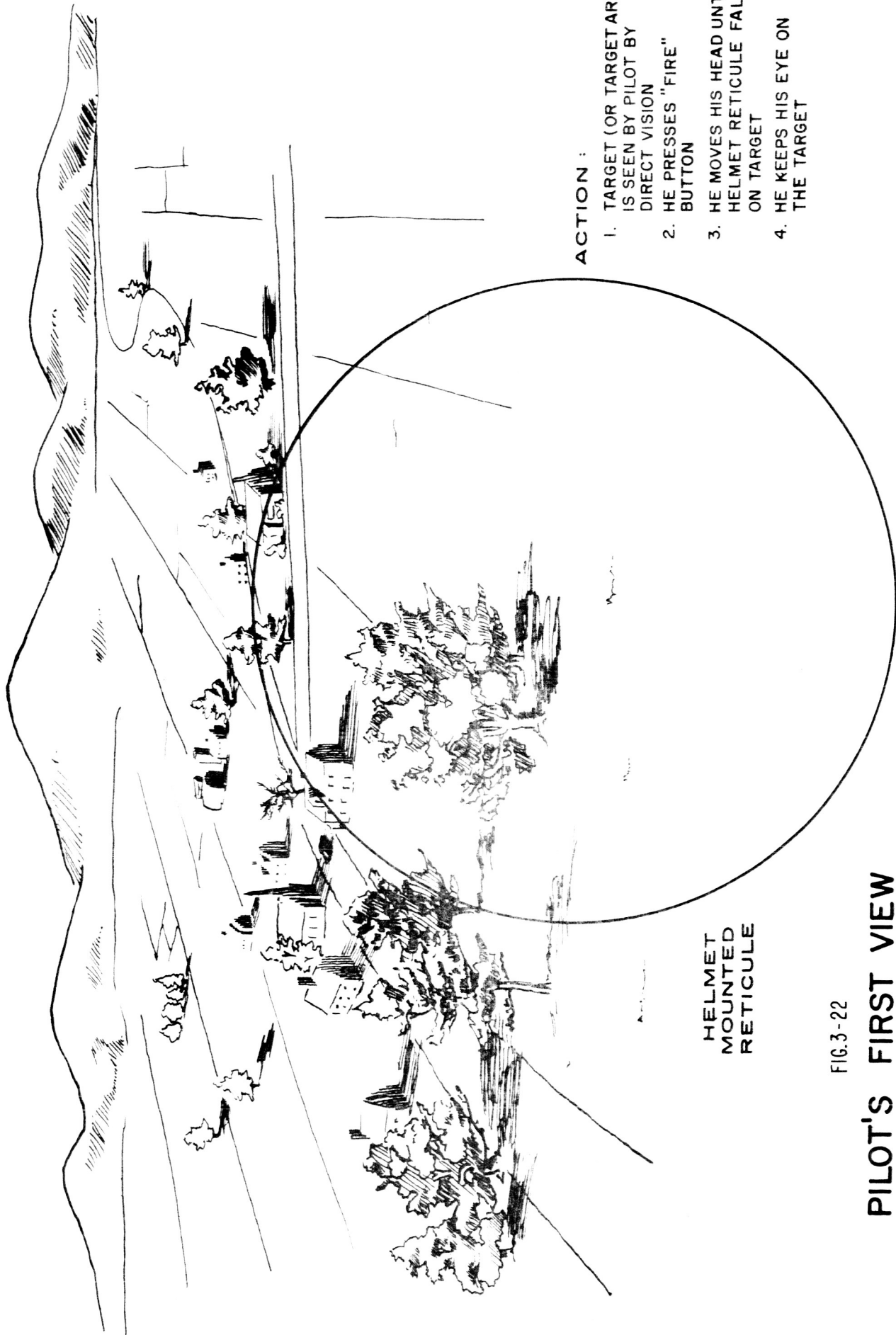
In the act of perceiving the target the pilot will naturally, and quite unconsciously, orient his eyes so that the image of the target falls over the acute part of his retina (fovea centralis). The oculometer will sense this eye motion and cause a command to be sent to the steerable optical system (G) in the missile. This will reorient the TV axis so that the target detail will be made to fall over the center of the TV system (e. g. , Figure 3-25). The oculometer will constantly monitor the pilot's eye position and cause whatever detail he chooses to look at to fall over the center of the TV system. To do this the steerable TV axis in the missile will be constantly directed at the target, and thus, will constitute a directing line for the missile-borne guidance system. As can be seen from Figure 3-26, the flight path error angle θ (i. e. , the angle between the velocity vector and the line joining the target and the missile) is given (in the simple illustrative two dimensional case) by:

$$\theta = \psi - \alpha$$

where α is the angle of attack and ψ is the angular deviation of the TV axis from the missile axis. Thus, by measurement (or estimation) of the angle of attack, α , an error angle can be deduced in the missile, from which optimum actuator commands can be computed.

During the flight of the missile to the target, the pilot's eyes will be used as the target sensing device and will not therefore be available all the time for direct vision. However, for most of the flight time of the missile the pilot would be able to flick his eyes momentarily away from the TV screen in order to monitor the flight situation of his own aircraft. These rapid flicks involve very high angular rate motions of the eye that can easily be distinguished from the relatively slow target tracking rates. A cut-out system can be employed to prevent these flicks from causing spurious motion of the camera axis in the missile.

Throughout the flight of the missile the pilot will be free to move his head in any way. Only a very small area of his visual field will be obstructed by the TV screen (Figure 3-24 and 3-25). Almost all of his peripheral vision will be free for direct vision. For many of the operations of flying an aircraft, peripheral vision may -- for short periods -- be adequate. It could, for example, be used as a safety monitor of a preprogrammed evasive maneuver to be executed immediately on firing the missile (Figure 3-27.).



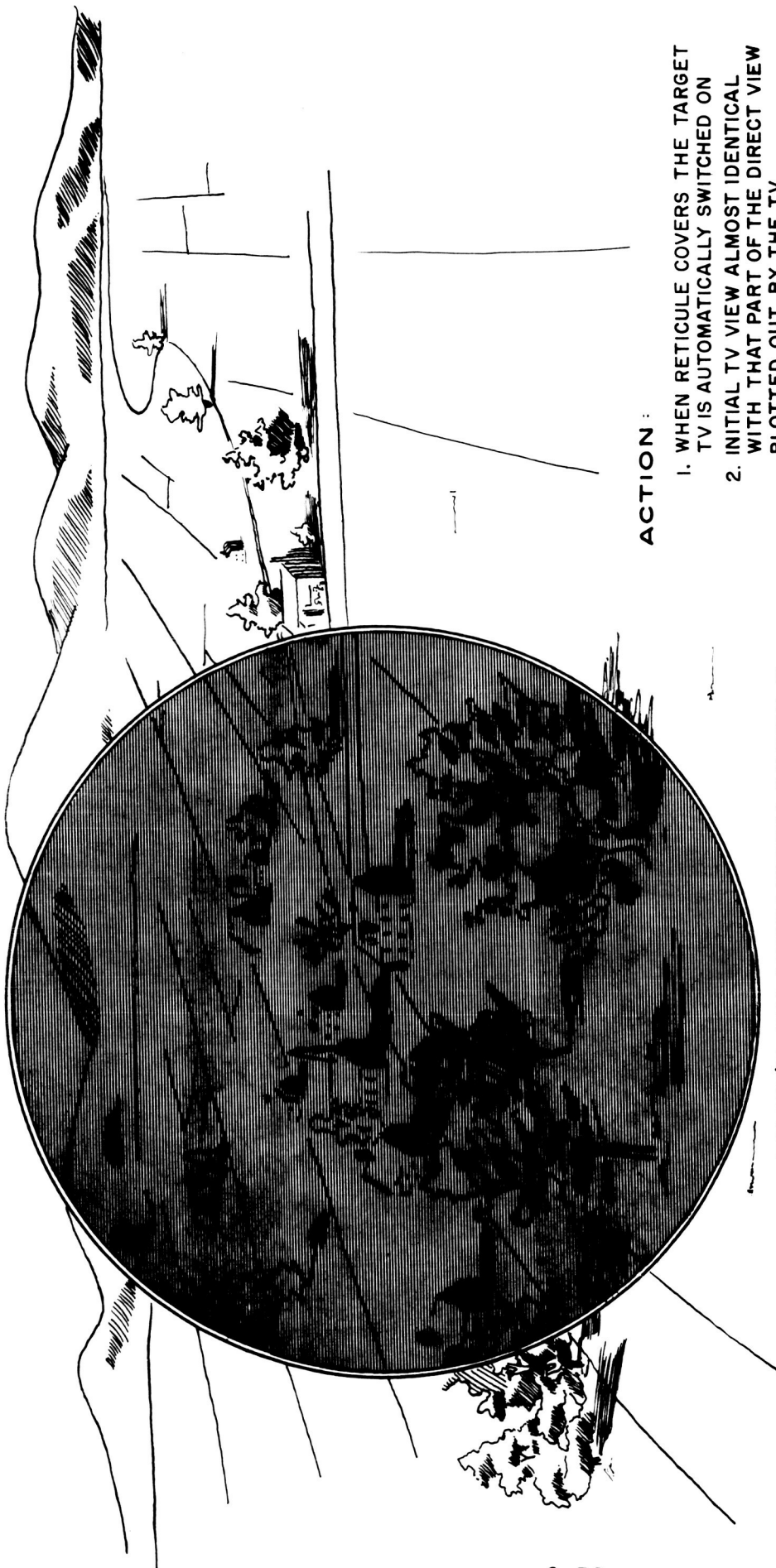
ACTION :

1. TARGET (OR TARGET AREA) IS SEEN BY PILOT BY DIRECT VISION
2. HE PRESSES "FIRE" BUTTON
3. HE MOVES HIS HEAD UNTIL HELMET RETICULE FALLS ON TARGET
4. HE KEEPS HIS EYE ON THE TARGET

HELMET
MOUNTED
RETICULE

FIG.3-22

PILOT'S FIRST VIEW



ACTION :

1. WHEN RETICULE COVERS THE TARGET
TV IS AUTOMATICALLY SWITCHED ON
2. INITIAL TV VIEW ALMOST IDENTICAL
WITH THAT PART OF THE DIRECT VIEW
BLOTTED OUT BY THE TV
3. MISSILE IS AUTOMATICALLY FIRED

PILOT'S TV VIEW FROM MISSILE

FIG. 3-23

MISSILE ABOUT TO BE FIRED

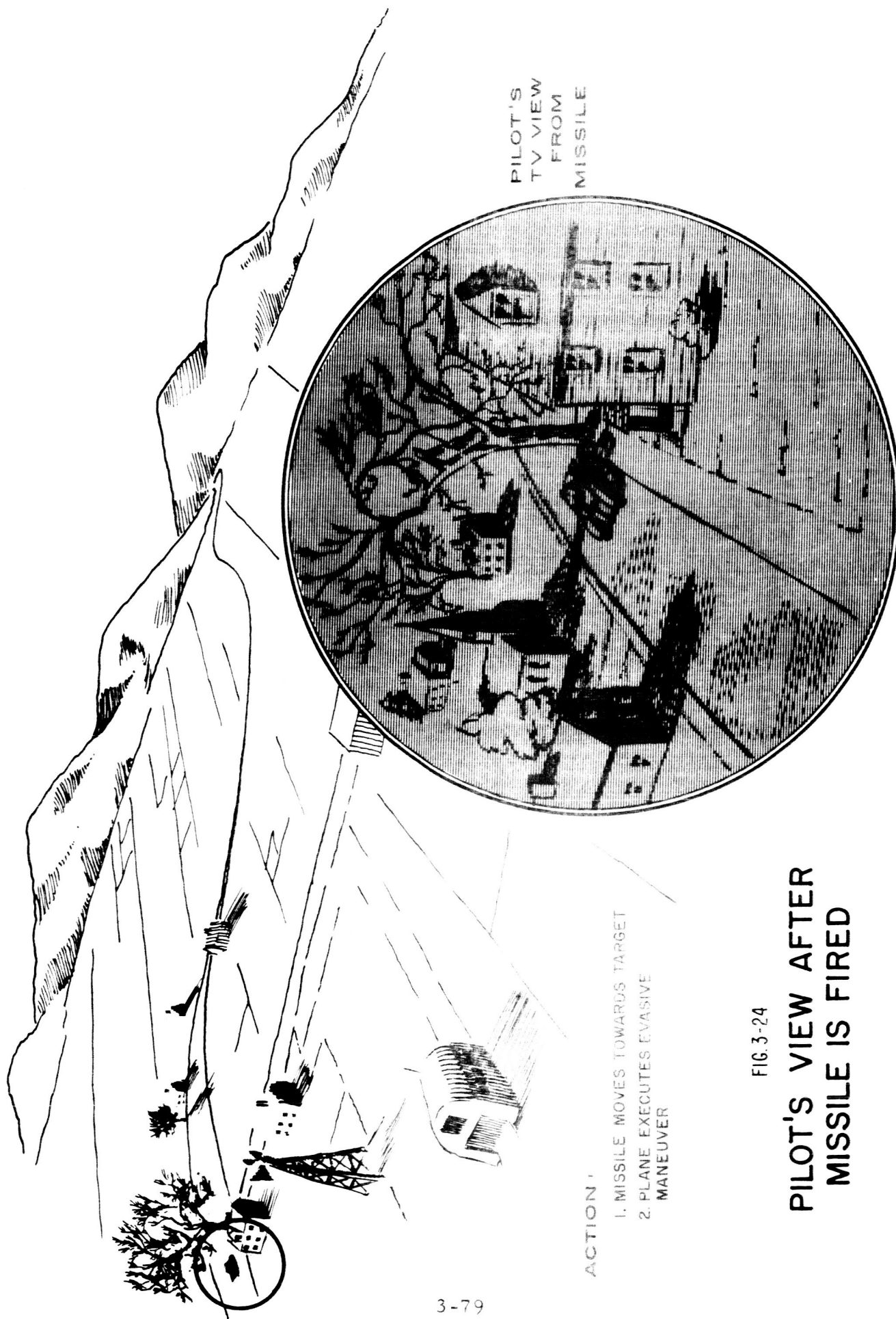
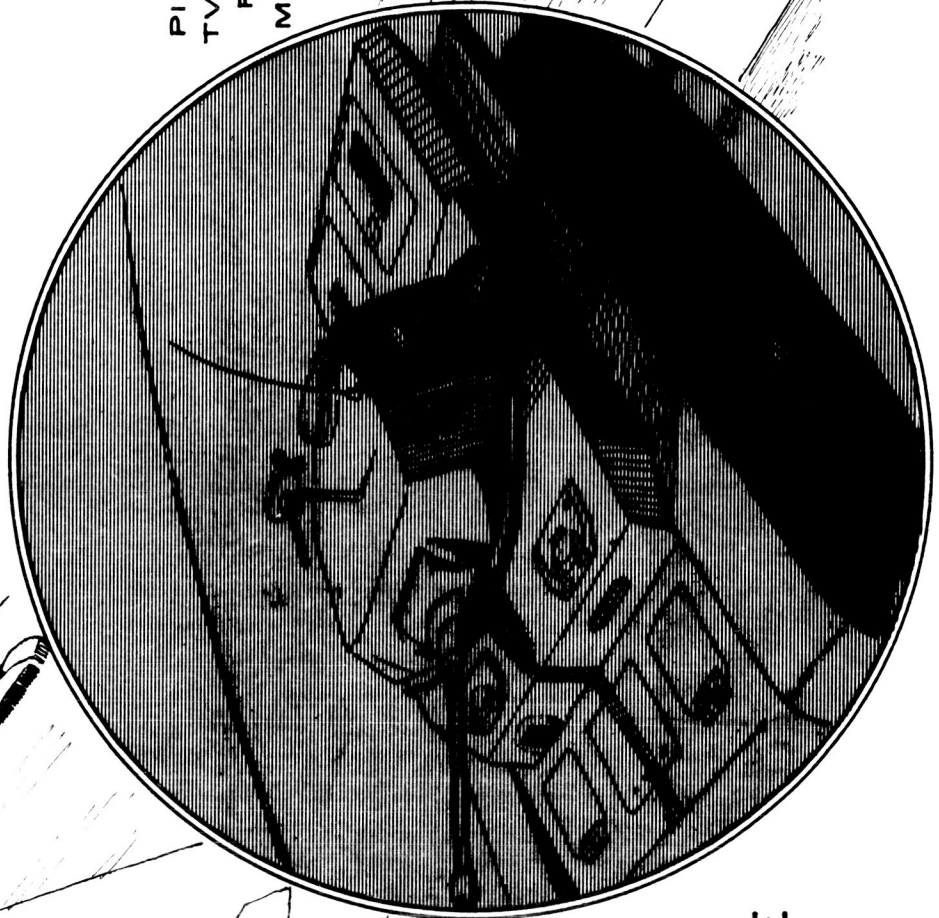


FIG. 3-24
PILOT'S VIEW AFTER
MISSILE IS FIRED

PILOT'S
TV VIEW
FROM
MISSILE



ACTION •

1. PLANE NOW FLYING WELL AWAY
FROM TARGET AREA
2. MISSILE VERY CLOSE TO TARGET

FIG. 3-25

**PILOT'S VIEW JUST BEFORE
MISSILE HITS TARGET**

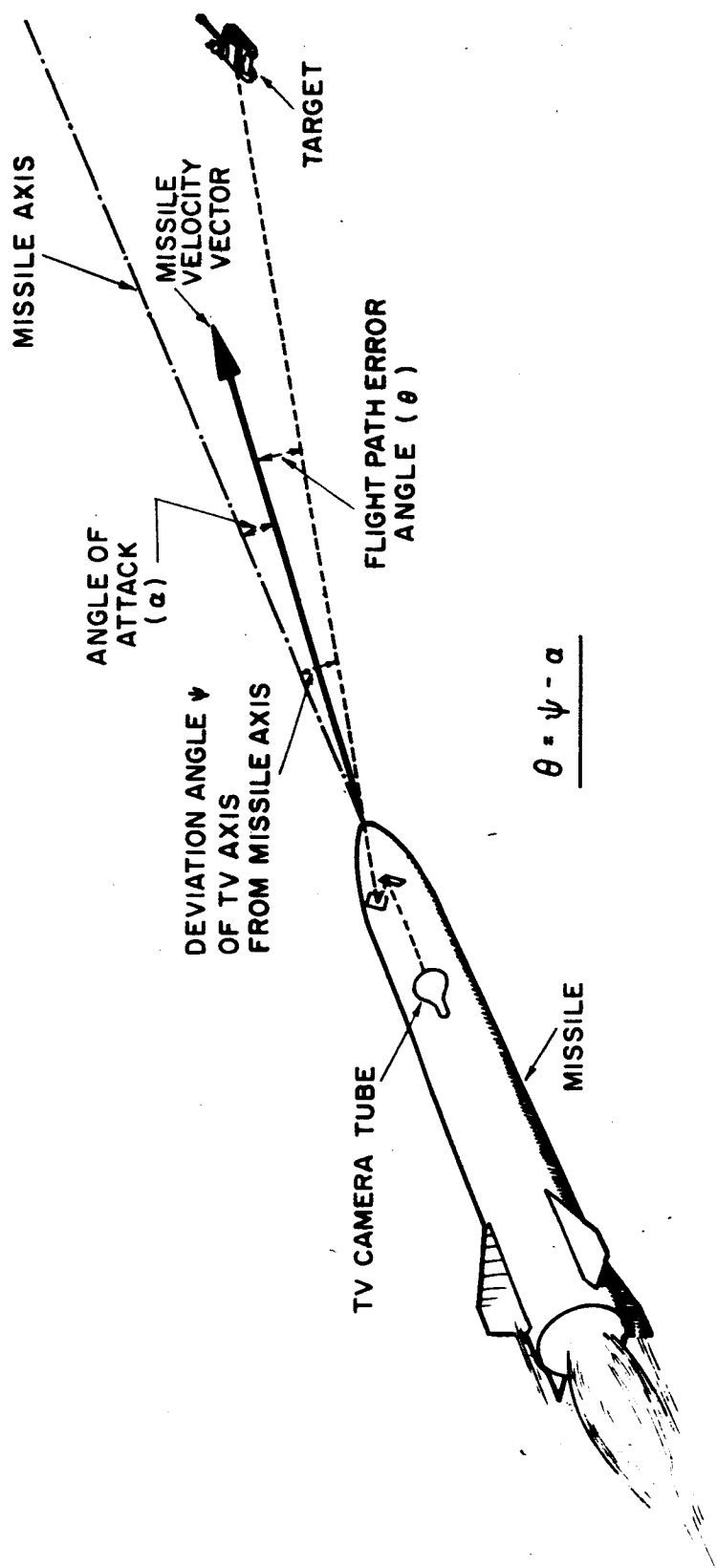


FIG. 3-26

GUIDANCE EQUATION

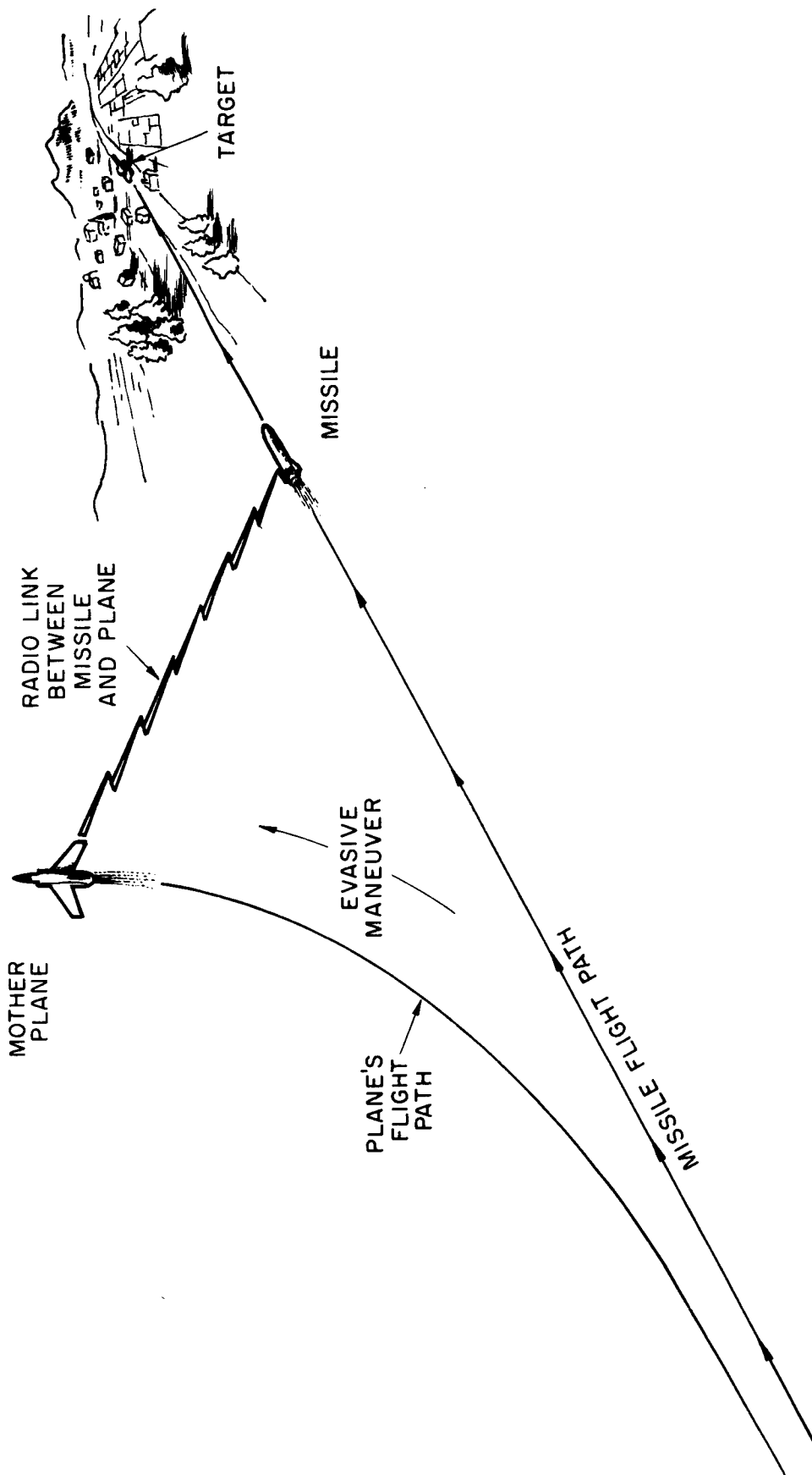


FIG. 3-27

EVASIVE MANEUVER AFTER FIRING MISSILE

B. Remote-Eye Vision - An Enhancement of the Visual Sense

Remote-Eye Vision is a TV system designed to provide a more complete visual sensation than ordinary TV. In its ultimate form, it would yield a visual sensation indistinguishable from reality. The subjective effect would be identical to that derived with the eye located at the camera - thus the name Remote-Eye Vision. This enhancement of the visual sense extends to remote distances, the full visual capability of man.

The primary limitation of ordinary TV is that under normal viewing conditions the TV image covers a very small part of the retina - approximately only 1/3 percent of the total area. Because most of the sense organ is not activated, ordinary TV fails to provide a sensory effect that even approximates "remote experience." This is in contrast to sound transmission systems, in which the sensory requirements of the ear can be so well satisfied as to render it sometimes impossible, for example, to distinguish between recorded music and a live performance. Since ordinary TV requires a bandwidth of the order of 5 mc and the TV image covers only 1/3 percent of the retina, it might appear that any significant increase in size of the TV image would be infeasible due to excessive bandwidth requirements. In proportion, a picture of an angular width of 2π steradians would require approximately a 1.5 kmc bandwidth. Yet it is the limitation of field of view that prevents presently existing TV systems from functioning as true extensions of the visual sense.

Thus it is clear that in order to realize the new dimensions implicit in Remote-Eye Vision the essential problem to be solved is how to provide, in a practical way, the full field of view that can be utilized by the eye. Without a wide field, TV is not a substitute for direct vision - with it, the exciting possibilities of complete duplication, or extension, of the human visual sense are opened up.

The impression we have that our eye presents us with a large detailed picture is an illusion. When the eye is fixed steadily at a single point only a very vague impression is gained of the visual scene at other points. The physiological reason for this is that the retina is not uniformly acute over its whole area. It is only in the very small foveal region that it is capable of perceiving detail. In the normal operation of the visual sense the eye muscles must rotate the eye so that the image of the detail of interest falls over the fovea.

At the fovea, the photoreceptors are mostly cones, packed very closely, each with a "private line" optic nerve fiber leading to the brain. Away from the fovea the total number of photoreceptors falls off, and also the proportion of rods to cones increases (see Fig. 3-28). In the peripheral areas of the retina groups of photoreceptors are connected "in parallel" through various synaptic connections in the retina itself. This is illustrated in Figure 3-29. Thus it is clear that the information flux to the brain, per unit area of retina, varies very considerably over the surface of the retina. It is highest at the fovea and lowest in the far peripheral part of the retina.

It becomes clear now that the very problem that seems to make Remote-Eye Vision infeasible - namely how to transmit the vast amount of electrical and optical information in a wide angle, detailed picture - exists for the eye as well. Nature has solved the problem by very sensibly avoiding it. There is no need for a wide angle picture of uniformly high quality. The brain (presumably) would be incapable of processing or utilizing the vast amount of information involved. Accordingly, the eye is constructed both "electronically" (i. e. in the retina) and optically to transmit to the brain a wide angle picture of variable quality. This makes the eye "feasible" - i. e. the simple 180° lens system of the eye does not have to give sharp undistorted images over its whole field of view. Thus the optic nerve trunk is kept to a reasonable size (a one minute of arc acuity over the whole visual field would require an optic nerve trunk as thick as the eye itself). A similar approach makes Remote-Eye Vision feasible.

In conventional TV no attempt is made to engineer an efficient match at the interface between man and machine. TV design has, formerly, simply involved the generation of an external image. Yet from the systems point of view the objective should be to inject a picture into the human nervous system, i. e. the TV image should be projected onto the retina in such a way that its information content is at all times matched to the perceptive capability of the part of the retina on which it falls. To do this the orientation of the eyeball must be monitored by the TV system. Thus a vital component of Remote-Eye Vision Television - which is of necessity a one viewer per camera system - is an oculometer to measure the orientation of the axis of regard of the viewer. The television image will be of variable quality. The highest quality part of the picture will always be projected on to the fovea.

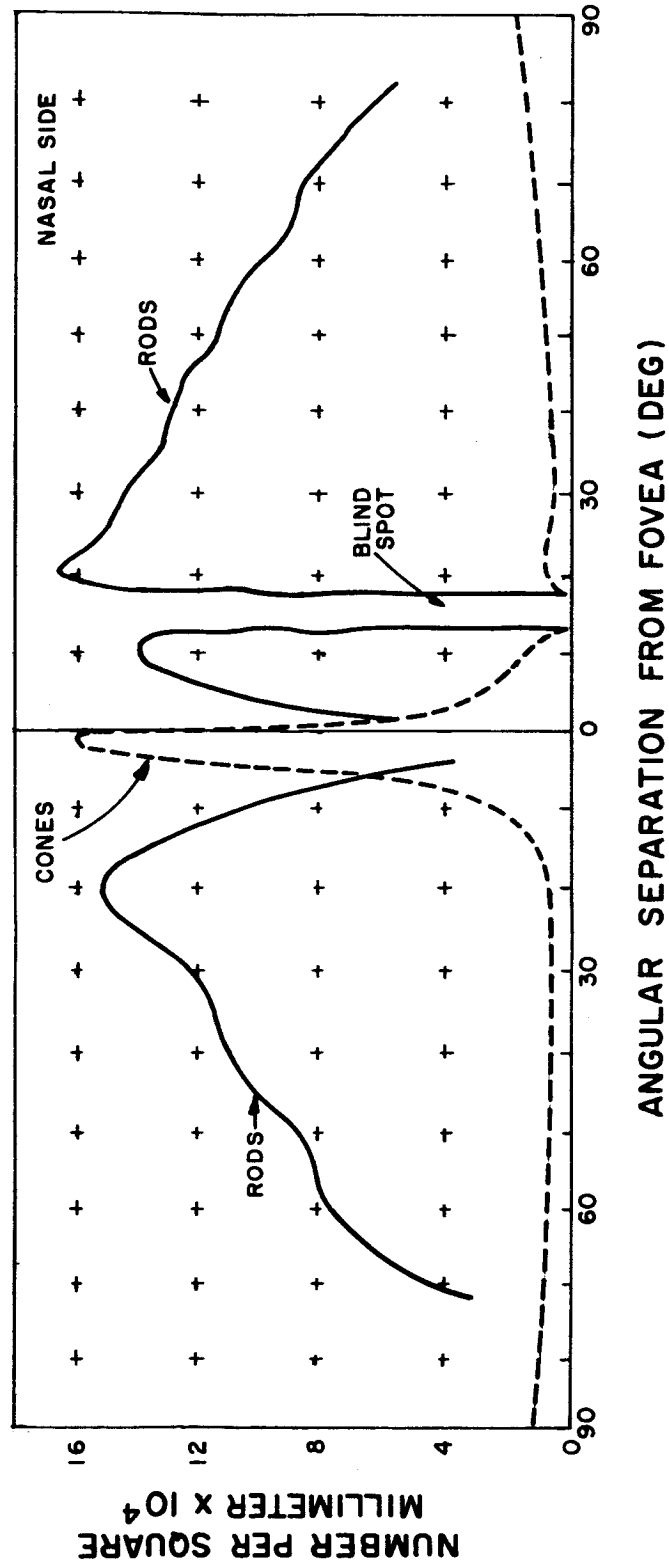


FIG. 3-28

DISTRIBUTION OF RODS AND CONES IN THE HUMAN RETINA

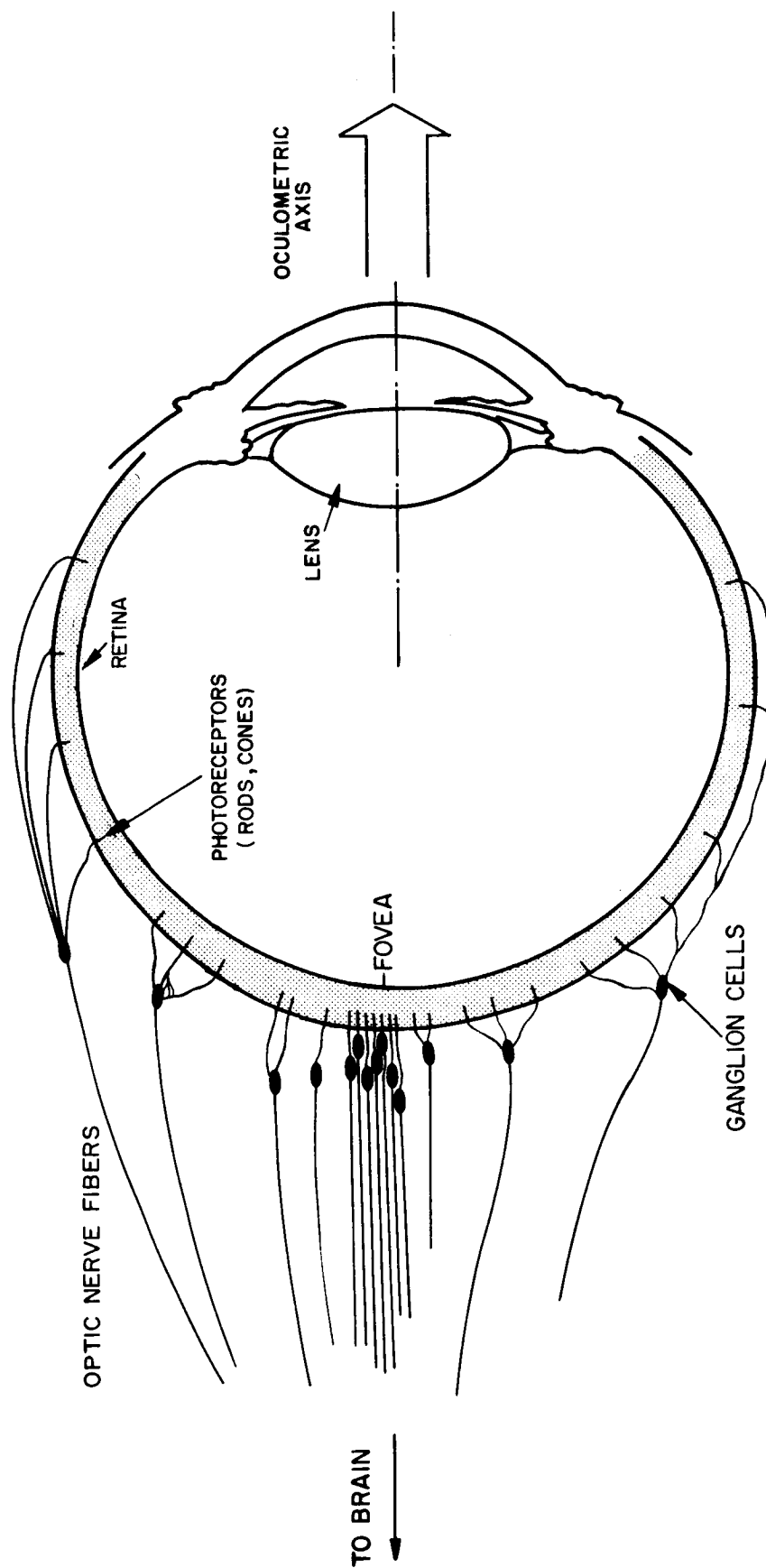


FIG. 3-29

SIMPLIFIED SCHEMATIC OF THE RETINA - BRAIN CONNECTION

Figure 3-30 shows the general arrangement of a variable quality TV system. The TV camera, monitor, and associated equipment are arranged so that at all times the center of the image plane of the camera corresponds to the center of the monitor image plane and the quality of the picture (e. g. resolution) is a function only of the angular distance from this central point.

Figure 3-31 illustrates the form the TV image would take. The oculometer monitors the position of the eye of the viewer and orients the variable optical axis unit - e. g. a gimballed mirror - (VOA1) in such a way that the high quality part of the image (at the center of the monitor tube) always falls over the fovea. It also controls VOA2 so that at all times the center of the TV camera tube is illuminated by that part of the scene that the viewer wishes to fixate.

In this system the eye is free to scan over the whole angular field of view. No matter in which position it settles, the TV system always arranges itself to transmit to the eye a picture, the quality of which is everywhere matched to the perceptive capability of the retina on which it is made to fall.

Note that the TV system itself - e. g. the surface of the picture tube - is stabilized on the retina. The scene being viewed is, however, not stabilized, and its image will move over the retina in a normal fashion.

The nature of the peripheral picture degradation that would be used is of great significance. The whole point of the system is that high quality can be achieved, with bandwidth economy, when the peripheral retina is supplied with information in a form it can accept. Thus development of the ultimate form of the system will require exact knowledge of peripheral retinal sensing in order that maximum bandwidth economy can be realized with a completely realistic display.

Relatively little is known about the sensing characteristics of the peripheral retina. Acuity, defined as the ability to resolve a grating, has been measured as a function of angle from the fovea. However, although acuity (together with the liminal brightness discrimination function) completely defines foveal vision, peripheral vision cannot be described in this simple way. Thus, for example, a grating subtending one minute of arc can just be resolved by the fovea. Alphabetic test letters are just identifiable

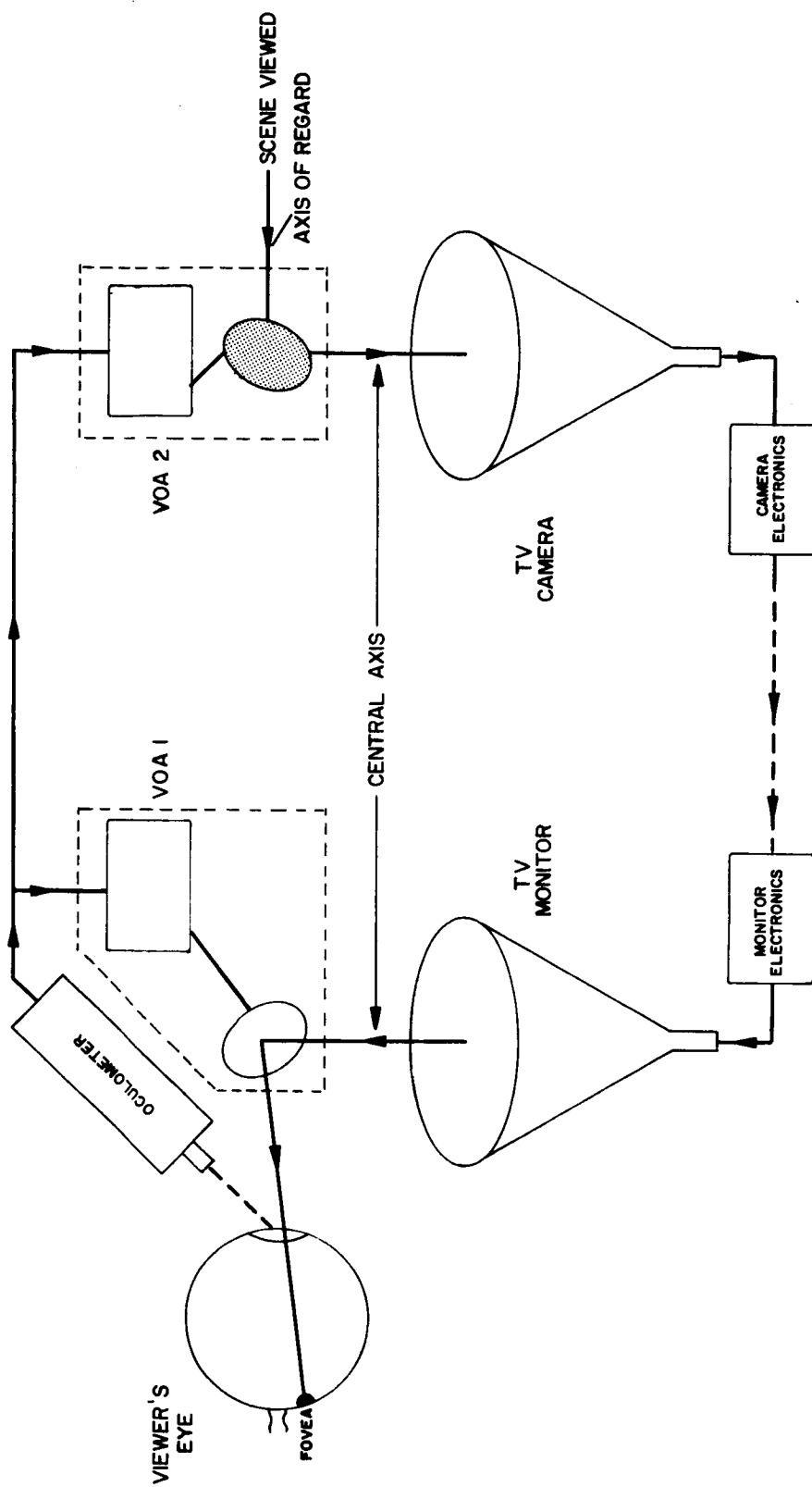
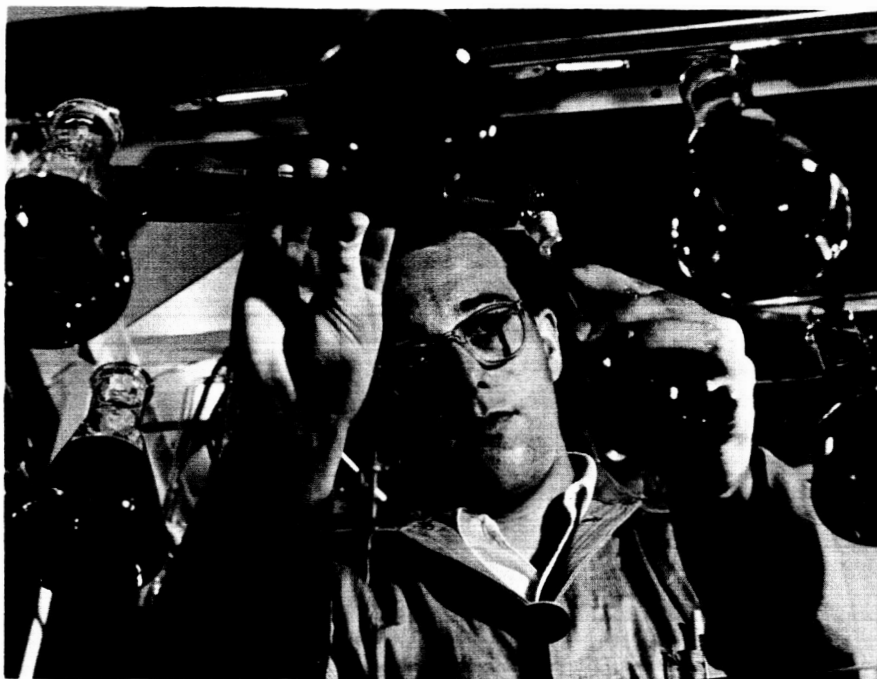


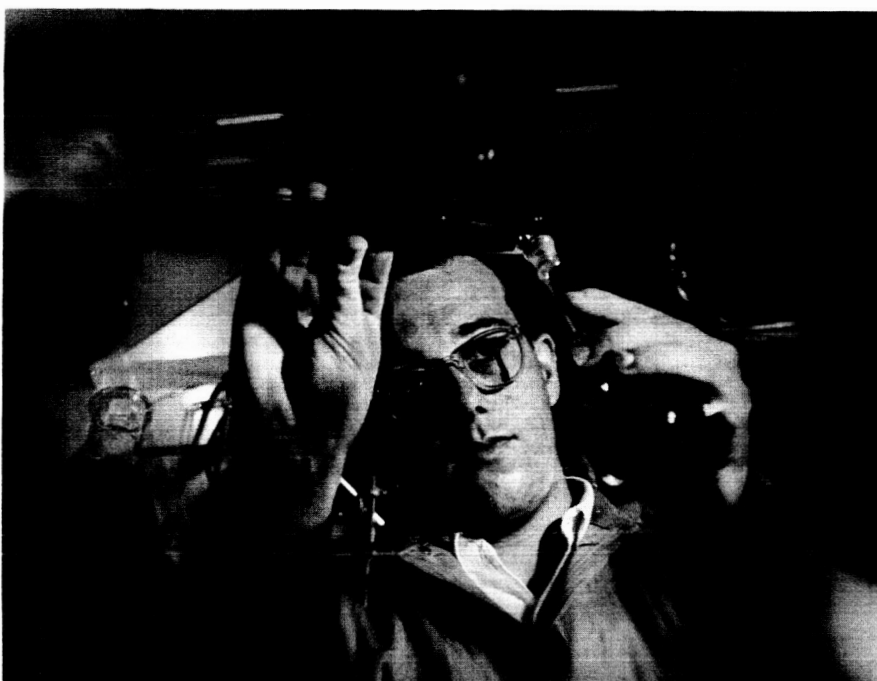
FIG. 3-30

TELEVISION SYSTEM

EXAMPLE OF A VARIABLE DEFINITION PICTURE



ORIGINAL WITH UNIFORM DEFINITION



VARIABLE DEFINITION

when they subtend about 5 minutes of arc. At a point in the periphery where the grating acuity has fallen to 10 minutes of arc it might be thought that, in proportion, a test letter would be resolvable when it subtended 50 minutes of arc. It will be found however that at such a point the letter must subtend many times this angle to be resolvable. Letter-resolving acuity falls off much quicker than grating acuity, as a function of distance away from the fovea. Peripheral vision is thus not just a less acute form of foveal vision. It is possible that it functions more as a sensor of the spatial-spectral content of the image rather than of the image itself. This is analogous to the ear, which senses not the actual pressure/time function at the ear, but rather the spectral content of this function.

A first approach to Remote-Eye Vision, employing lower resolution in the peripheral TV image, cannot be expected to result in a perfect match between the TV system and the retina. Nevertheless this simple first approach could constitute a substantial improvement over ordinary TV, where no attempt at all is made to match the TV system to the variable characteristics of the retina.

The results of the evaluation of an initial Remote-Eye Vision system would throw considerable light onto the exact nature of peripheral vision - in a way that could not be done by any other method.

System Application - Remote Interaction

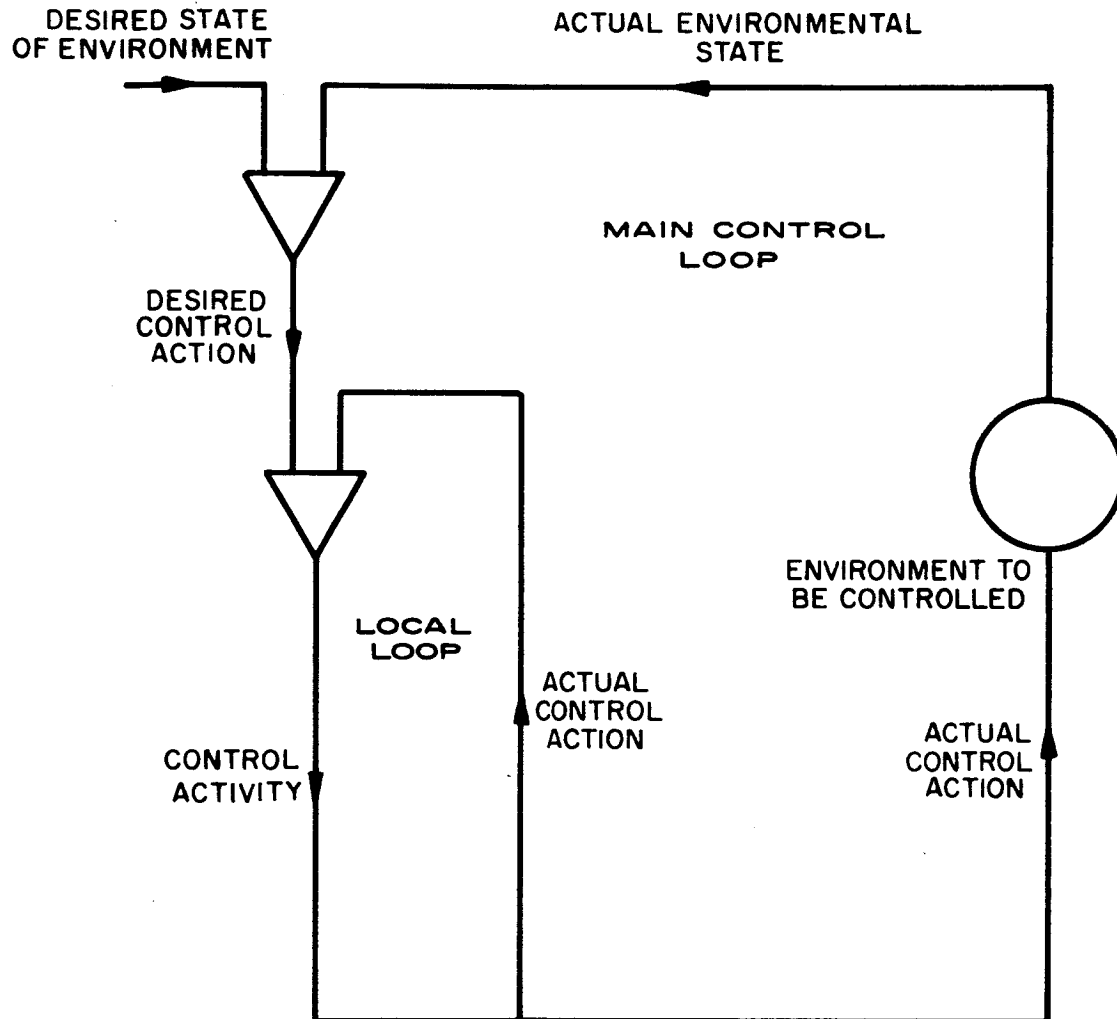
Man's normal interaction with his environment may be regarded as a closed loop information exchange process, as illustrated in Figure 3-1. Man absorbs information from the environment via his senses, processes it, and then - according to his preconceived goals or will - exerts an effect on the environment by means of his motor capabilities. This effect may be by hand, foot, speech, facial expression, etc. Essentially the 'effect' is a return to the environment of information from the man.

The to-and-fro information flux can be recognized within the body itself where the loop is closed by the brain - which is functionally connected to the outside world solely by data links.

The analogy between man's interaction with his environment and a typical feedback control system (as illustrated in Fig. 3-32) has been recognized for some time. The three basic elements of each system are:

FIG.3 -32

TYPICAL INANIMATE FEEDBACK CONTROL SYSTEM



1. information sensing devices
2. motor devices
3. central data processing system

In the layout design of inanimate feedback control systems, the central data processing device is located wherever is most convenient. If there are special difficulties associated with its location close to the environment which is being controlled, it may be located remotely, and connected to the sensing and motor devices by information transmission lines.

Analogously, man may wish to interact with a remote environment when there are special difficulties associated with his actual physical location at the environment. He may do so by mail, or by telephone, (both potentially information exchange processes) or even by a combination of television and remote manipulators.

Two very important limitations of most remote control arrangements are noted:

1. The remote sensory and motor devices are very inefficient, or are coupled very inefficiently to man's sensory and motor organs. For example, television is very inadequate as a remote extension of the visual sense. Less than 1/2% of the total retinal surface is activated by the image of a 500 line television picture, under normal viewing conditions.
2. In most remote control situations an extra control loop is involved. The man is in contact with both his local environment, and the remote environment.

The result of the first limitation is that the effective data rate between the man and the remote environment is inevitably less than in the case of the normal interaction of a man with his environment. With television, for example, only a very small fraction of the data capacity of the visual sensory channel to the brain is used to convey information from the remote environment.

The second limitation means that the brain does not operate in its normal mode in relation to the remote environment. When the central data processing unit in an inanimate feedback control system is located remotely, the form of the loop is not changed. The control unit receives data only from the remote sensor and can actuate only the remote motor device.

Thus, for such inanimate control systems, remote operation can be made virtually indistinguishable from non-remote-operation. In the conventional human remote control arrangement, on the other hand, the brain receives data both from the remote and local environments, and generates motor activity directly, only in the local environment. Thus quite apart from any limitation that may exist in the data links, the brain cannot operate in relation to the remote environment, as it would were the man actually located in that environment.

In ordinary remote control systems therefore man does not have a true "remote presence" capability because he is mainly in contact with his local environment.

It is the purpose of this section to draw attention to the theoretical and practical significance of the concept of Remote Interaction - a system that may allow a man to function in a remote environment without actually being there.

Remote Interaction is defined as an ideal system in which zero delay, distortionless information links (including sensors and transducers) are provided for all of the human sensory and motor contacts between a man and a remote environment. The arrangement of the sensors and transducers at the remote location could be in replica of the human form.

A man using such a system would, by definition, suffer a subjective experience of the remote environment identical to that of normal on-the-spot contact with the environment. Furthermore, his ability to perform any task in the remote environment would be judged, objectively, as being identical to his normal on-the-spot functional capability. It may be noted that in Remote Interaction, both the limitations of conventional remote operations, discussed earlier, are absent.

The purpose of going to the moon or planets is not to locate the human body at these places but to place a human into normal contact with these environments. At present this can only be done by travel. Thus space travel has become synonymous with our space goals. Remote Interaction, by virtue of its definition, is a completely equivalent alternative to travel and for this reason is, at least, of theoretical significance. Travel means relocation of the body, and thus of the sensory and motor organs. Proved that distortionless extensions of the senses and motor organs are available, the

total effect of body relocation can be obtained without transportation of the body.

An ideal Remote Interaction system is clearly impractical. It might also be thought that the best possible practical approach to it would be a very expensive system, very difficult to make, that would be so inferior to the theoretical ideal as to be of little value.

This may not be so however. In the first place, very little serious engineering effort has been devoted to the technological problems of providing efficient remote extensions of the human senses and motor organs. High fidelity sound probably represent the only well developed sense (hearing) and motor (speech) remote extension system. It is suggested that there are no insuperable problems involved in making efficient extensions of most of the other important senses and motor organs. The data rates of the senses appear to be within the state-of-the-art. Likewise the important motor systems of the body involve only a relatively small number of pivot points and degrees of freedom. Although construction of an adequate Remote Interaction system might well prove to be a major undertaking, there does not appear to be any fundamental physical limitation involved that could make it impossible. The approach to be followed, in developing the various sensory and motor extensions, would certainly involve an engineering analysis of the corresponding human systems, in order that the maximum possible coupling efficiency be attained. As will be described later, probably the most difficult sensory extension system to construct - remote vision - becomes, at least, potentially feasible when the system structure of the human visual mechanism is understood.

Secondly, the magnitude of the development program needed to construct a Remote Interaction system must be judged in relation to the size of the total space program. Any method of space exploration seemed infeasible to many people only a few years ago. The conventional method now being pursued, of transporting a man to the remote location and bringing him back again, itself requires a prodigious technological effort. Should it be possible therefore to use Remote Interaction to augment manned operations, development of a Remote Interaction system might well prove to be economical.

Thirdly, the limitations of a practical Remote Interaction system - which must necessarily be inferior to the idealized conceptual system - should

be judged against the very real limitations of direct, manned exploration. It must be recognized that man will never be able to explore the moon or planets in the same sense that our ancestors explored the jungles of Africa or the polar regions. He will always have to wear a space suit of some form and the full range of his sensory and motor capabilities can, therefore, never be brought into natural contact with all these extraterrestrial environments. Remote Interaction could actually prove to be a more 'natural' way of exploring certain alien environments, since the sensory and motor organs may thereby be brought into contact with the environment more effectively than would be possible directly through a cumbersome life supporting space suit. There are some environments in which man may never be able to locate his body no matter how well protected he may be.

The practical significance of the concept of Remote Interaction is, therefore, that construction of a practical Remote Interaction system - although a major undertaking - appears to be technically feasible; the very special difficulties and inherent limitations of manned exploration of extraterrestrial environment suggest also that it might be worthwhile.

As a first step, towards practical implementation of the concept, the overall requirements of the sensory and motor links and transducers may be determined for a remote Robot system adequate to either:

1. perform a very narrow range of tasks, e.g. to use a wrench,
2. perform a range of tasks, e.g. to assemble a space station,
3. provide a wide range of human skills and sensory capabilities as might be required to explore and exploit an unknown environment such as the Martian surface.

From the specifications of the necessary transmission links and transducers, an estimate can be made of the practical feasibility of this approach. The advantages and disadvantages of Remote Interaction in terms of functional performance and cost, can be balanced against the corresponding advantages and disadvantages of the conventional method of physically transporting an astronaut into the remote environment.

A physical implementation of the Remote Interaction Concept would consist of three main elements - coupling devices to be fitted to the senses and motor limbs of the man, an assembly of sensors and mechanical actuators at the remote location, and a two-way transmission link between the

man and the remote environment.

The sensors and actuators in the remote environment might be assembled into a configuration that would look like a robot in outward physical appearance. It is proposed, however, to introduce a special word - Humanoid - to describe the remote part of the system. The word robot, as generally understood, means a mechanical 'man' with its own (machine) will and intelligence that can function independently of direct human control. The system of sensors and actuators that would be deployed at the remote location as part of a Remote Interaction system, is clearly not a robot as defined above, since it has no intelligence of its own, but is simply a remote extension of the functional capabilities of a man. Whereas both a Humanoid and a robot might look and behave like a mechanical man, a robot is controlled by its own inanimate, preprogrammed computer, a Humanoid will be controlled by a human mind.

The Humanoid will have sensory detectors corresponding to all the principal senses - e.g. vision, hearing, finger touch, etc., but excluding the myriad touch and pain sensors all over the body that, while essential for the maintenance of the body, are not really involved to any significant extent in man's interaction with his environment. The Humanoid sensory detectors would be coupled to the senses of the man controlling it. Thus the man would have a "Humanoid's-eye-view" of its environment. He would see nothing of his own local environment - a remote visual capability far better than ordinary TV is postulated. The muscular action of the man, by hand, arm, finger, etc., would be monitored and made to command functionally identical Humanoid motion. As the man moved his arm and fingers, he would feel his own arm move (kinesthetic sense) but it would be the Humanoid limbs that he would see moving and they would appear to fall within his field of view just where he would expect to see his own limbs were his eyes free to monitor the local environment. If the Humanoid's hand should come into contact with an object in its environment, the man would experience a sensation of touch which would correlate well with his visual sense showing the Humanoid's hand touching the object, and also his kinesthetic sense of hand position. In this way a large fraction of the significant sensory cues would be duplicated. The number, form and kind of data inputs would be engineered into the human system; the human data processing and control mechanism (i. e. the brain) would be

operating in its normal mode. The man would - to a large extent - be in effective contact with the remote environment through the body of the remotely located Humanoid: he would feel himself to be "in" the remote environment.

The various subsystems of this system are considered below:

SENSES

Vision

Vision will generally be the most important sense - in the present context - because it represents the largest single sensory channel capacity into the brain.

The inadequacy of ordinary television as a remote extension of the visual sense is evident both physiologically and by practical experience. Only 1/3% of the retinal surface is covered by a 500 line television picture under normal viewing conditions. Thus most of the 10^6 optic nerve fibers that link the retina to the brain are not used. The extreme impairment of general visual capability imposed by the narrow field of an ordinary TV system, can be readily appreciated by artificially constricting the field of the eye from the normal value of about 180° to the same size as a TV screen at normal viewing distance - i. e. about 9° . It will be found that functional performance in carrying out many everyday tasks is then very adversely affected.

In spite of the apparent inadequacy of television, the actual data flux is high. Thus in domestic television, approximately 7×10^6 resolvable picture elements are displayed per second. The actual data rate to the brain from the retina must also be of this order, since there are only about 10^6 optic nerve fibers and the reaction time of the eye is between 1/10 and 1/100 second. Thus it appears that the inadequacy of television is not due to a limited data flux capacity, but rather to the fact that the television picture is not matched to the retina. It is of uniform quality (in resolvable elements per unit area) whereas the perceptive capability of the retina is very non-uniform. The use of ordinary television as a remote viewing device in the proposed system appears to represent an extremely poor man/machine interface design.

A television system for the Humanoid should be constructed in which the television image itself is stabilized on the retina, independent of eye motion, with the optical axis of the camera slaved to the angular attitude of the viewer's eye. In this way the image of the remote scene would not be stabilized, but would move over the retinal surface in just the same way as if the eye were observing the remote scene directly. With the television image stabilized on the retina however, the video data flux could be tailored to the sensory requirements of each part of the retina - thus allowing a much greater area of the retina to be covered without unduly increasing the total data flux through the television system.

A television system of this type would require measurement of the direction of pointing of the eye of the viewer, with use of this information to control both the projection of the monitor image into the eye and the angular attitude of the camera axis. Experimental techniques are available for measuring eye motion and could be developed into a suitable operational form.

The optical axis of the remote television camera would move like a remote eyeball, constantly flicking over the viewed scene in response to the motion of the viewer's eye. Transmission delays between the man and the Humanoid would probably limit the range of this system to about 10,000 miles. At this range there would be a delay of about 1/10 second between a movement of the eye and a corresponding movement in the received picture.

It is recognized that not enough is known of the nature of peripheral vision to enable the most efficient match possible to be made between the retina and the television system. However, enough is known to enable a considerable improvement to be made over present television systems. Research, to find more exact qualitative and quantitative descriptions of peripheral vision, would undoubtedly be an important part of a program to fully develop this form of remote viewing.

Touch

From the point of view of providing remote manual dexterity, sense of finger touch would be an important part of the Humanoid system.

The remote Humanoid would be fitted with a number of touch sensors on its fingers, and these would be linked to corresponding touch simulators fitted to the fingers of the man controlling it. It is possible that a fairly small number of touch sensors, less than 100, would provide an adequate remote sense of touch for both hands.

Balance

To some extent this could be a "subconscious" closed loop within the Humanoid itself. This would tend to mitigate stability problems arising from the 1/10 second round trip communications delay at the 10,000 mile maximum range.

Hearing

This might be useful for certain applications and would obviously be a very easy sense to instrument.

MOTOR FUNCTIONS

Arm, Hand and Finger Motion

The man in the system would be fitted with sensing devices to monitor the position of his arms, hands and fingers. These signals would be transmitted to the Humanoid and used to command identical remote limb motion.

Feet Motion

As above, to the extent demanded by the actual system design. For many applications it might be preferable for the Humanoid to move on wheels rather than on feet. In this case the man might operate foot pedals which would control the motion of the vehicle directly.

Speech

This could easily be incorporated as necessary.

TABLE II

Task	Problem Involved	Present Solution	Proposed Solution	Advantages of Proposed Solution
1. Maintenance and repair of operational satellite systems; e.g., communications, meteorological, etc.	The cost of these systems is high. To be economic they must therefore have a reasonably long life - e.g., 5, 10 or 20 years	Long operational life can be attained by using components with very high reliability, limiting the size and complexity of such systems, replacing defective systems with the launch of a new satellite, etc.	Establishment of an orbiting repair facility with at least two Humanoids (for self-repair) having an inter-orbit maneuvering capability so that faulty satellites can be approached and then repaired by the Humanoid.	Increase in lifetime of all our operational satellite systems, thus enhancing their economic status. More complex systems can become economically feasible (in spite of a low mean time to failure).
2. In flight repair of manned spacecraft.	Under certain circumstances (e.g., high radiation flux) it may be very difficult or impossible for an astronaut to go outside the spacecraft to make external repairs.	Either use very heavy protective space suits or highly specialized remotely controlled tools.	External repairs to the spacecraft carried out by a general purpose Humanoid, which will be controlled by an astronaut from within the cabin.	External repairs can be carried out without any danger to the astronauts, and without the need of highly specialized tools.
3. Pilot for orbital military satellites, e.g., Dyna-Soar, etc.	Such systems will require either (a) the constant presence of a man in orbit or (b) rapid and reliable deployment of a man into orbit for short unpredictable periods.	Ferrying of astronauts to and from orbit.	Humanoid permanently stationed in orbit with controller located on earth.	Present methods are so difficult, expensive, impractical etc., as to render virtually infeasible the task of carrying out any sort of space patrol, on an operational basis.
4. Initial Lunar exploration.	Very large boosters required to give highly reliable safe return capability.	Wait for development of very large boosters i. e. probably until 1967-1970.	Soft land device portion of the system with power supplies, communications etc., on to surface of moon. Dispatch human controller to moon orbit (human controller could be located on earth, but the 2 1/2 sec. communication delay would considerably impair his capability.	Lunar exploration accomplished several years ahead of the first manned mission.

The first requirement indicates that the measurement should be optical in nature. The third indicates that the front surface of the eye, rather than the retina itself, should be viewed by the optical sensor. The fourth requirement then indicates that it should be the area within and on the pupil/iris boundary that should be tracked since this part of the eye is necessarily always exposed when the eyes are being used.

A unique feature of the new method, to be described, is that two points in the eye are located. In this way eye position can be measured independently of the lateral position of the instrument relative to the head and thus requirements (2) and (1) above may both be satisfied.

Description of Method

One of the simplest methods of eye position measurement is that of direct observation of the position of the pupil. The position of the pupil may be sensed by a special TV scanning system, acting as a pupil tracker (Fig. 3-33).

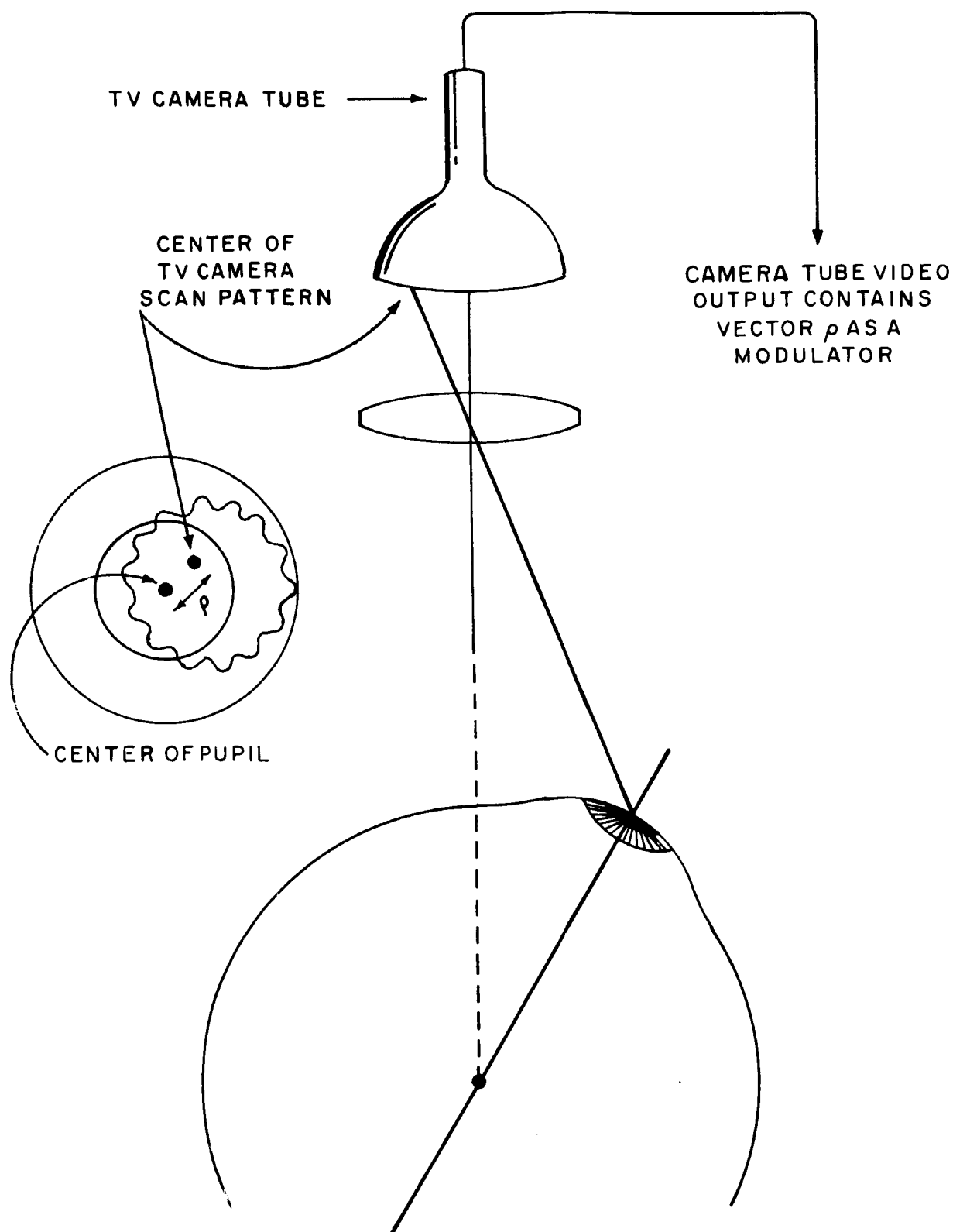
The position of the pupil, relative to the head, defines the orientation of the eyeball. However, any movement of the measuring device (i. e. the pupil tracker) relative to the head will cause large errors.

This new method of eye measurement does not require that the measuring device (i. e. oculometer) be rigidly fixed to the head. It involves measurement of the position of both the pupil and the center of curvature of the front corneal surface of the eye.

The front corneal surface of the eye acts as a concave mirror with a reflection coefficient of approximately 2.5 percent and radius of curvature 7.7 mm. The posterior surface of the cornea and other surfaces in the eye have much lower reflection factors, so that an easily identified, unique, image (Y) of an external point X is formed by reflection in the front surface. The external point, its (virtual) image by reflection, and the center (C) of curvature of the front corneal surface of the eye all lie on one straight line (Fig. 3-34).

The basic measurement method is illustrated in Figure 3-35. A scanning TV system acts as a pupil tracker. Associated with it, on the same optical axis, is a CRT display of a circle (X), the position (and radius) of which can be controlled electronically. The pupil tracker senses any deviation of its scan pattern from both the pupil and the image (by corneal

FIG. 3-33
SIMPLE METHOD OF OCULOMETRY



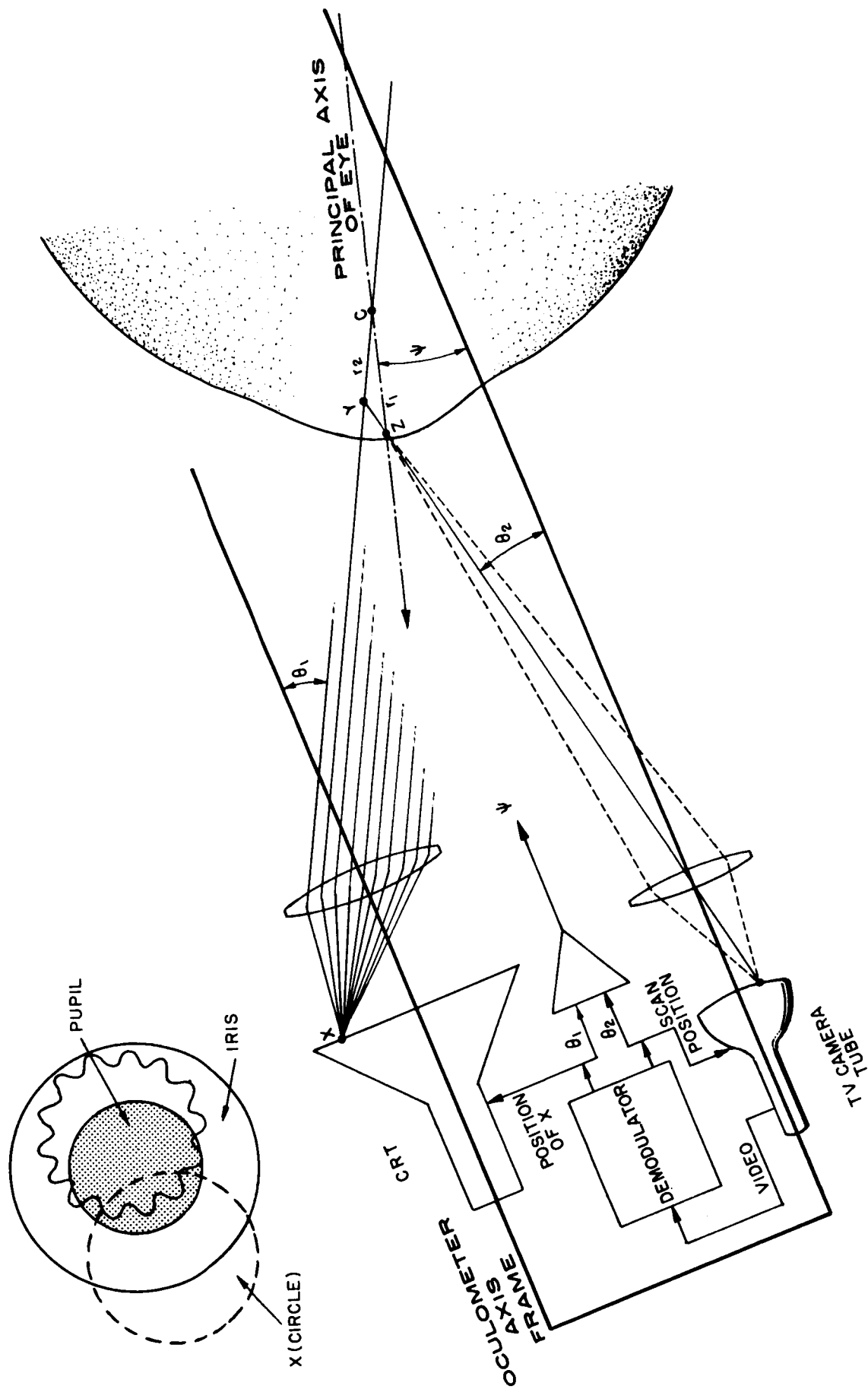


FIG. 3-35

MEASUREMENT METHOD

reflection) of the circle X. (It may be noted that the pupil appears, after refraction at the cornea, to be about 2.5 mm within the eye, and the image of X (when at infinity) to be about 3.8 mm within the eye.) Any deviation of the scan pattern from the pupil causes the scan pattern to be repositioned in a closed loop servo control action. Any deviation of the scan from the image of X causes the position of X to be changed in a second servo loop. (Not shown are other control actions i.e. adjustment of the diameter of the scan pattern and the amplitude of the sinusoidal component of the scan.)

When both the servo loops shown in Figure 3-35 have performed their function the image (Y) of X in the cornea, and the apparent center (Z) of the pupil will appear - to the pupil tracker - to be in line. The geometry under these conditions is as shown. Relevant detail is shown again in Figure 3-36. The distances YC (r_2) and ZC (r_1) are constants for any given eye. Applying the sine rule to the triangle ZYC;

$$\frac{r_1}{\sin(\theta_1 + \theta_2)} = \frac{r_2}{\sin(\theta_2 + \psi)}$$

where θ_1 θ_2 are the (electronically) measured bearing angles of the center of circle (X) on the CRT, and the position of the pupil scan. ψ is the angle between the principal axis of the eye and the optical axis of the oculometer. (The principal axis of the eye is defined here as the line joining the center of curvature of the front corneal surface and the apparent center of the pupil.)

Using small angle approximations

$$\begin{aligned}\psi &= \frac{r_2}{r_1} (\theta_1 + \theta_2) - \theta_2 \\ &= \frac{r_2 \theta_1 + \theta_2 (r_2 - r_1)}{r_1}\end{aligned}$$

Substituting nominal values (for a standard eye) of r_1 r_2 ,

$$\psi = \frac{3.8}{5.2} \theta_1 - \frac{1.4}{5.2} \theta_2$$

As shown in Figure 3-35 the desired angle ψ is computed, as above, from θ_1 and θ_2 .

Note that ψ is derived independently of the lateral position of the eye relative to the oculometer.

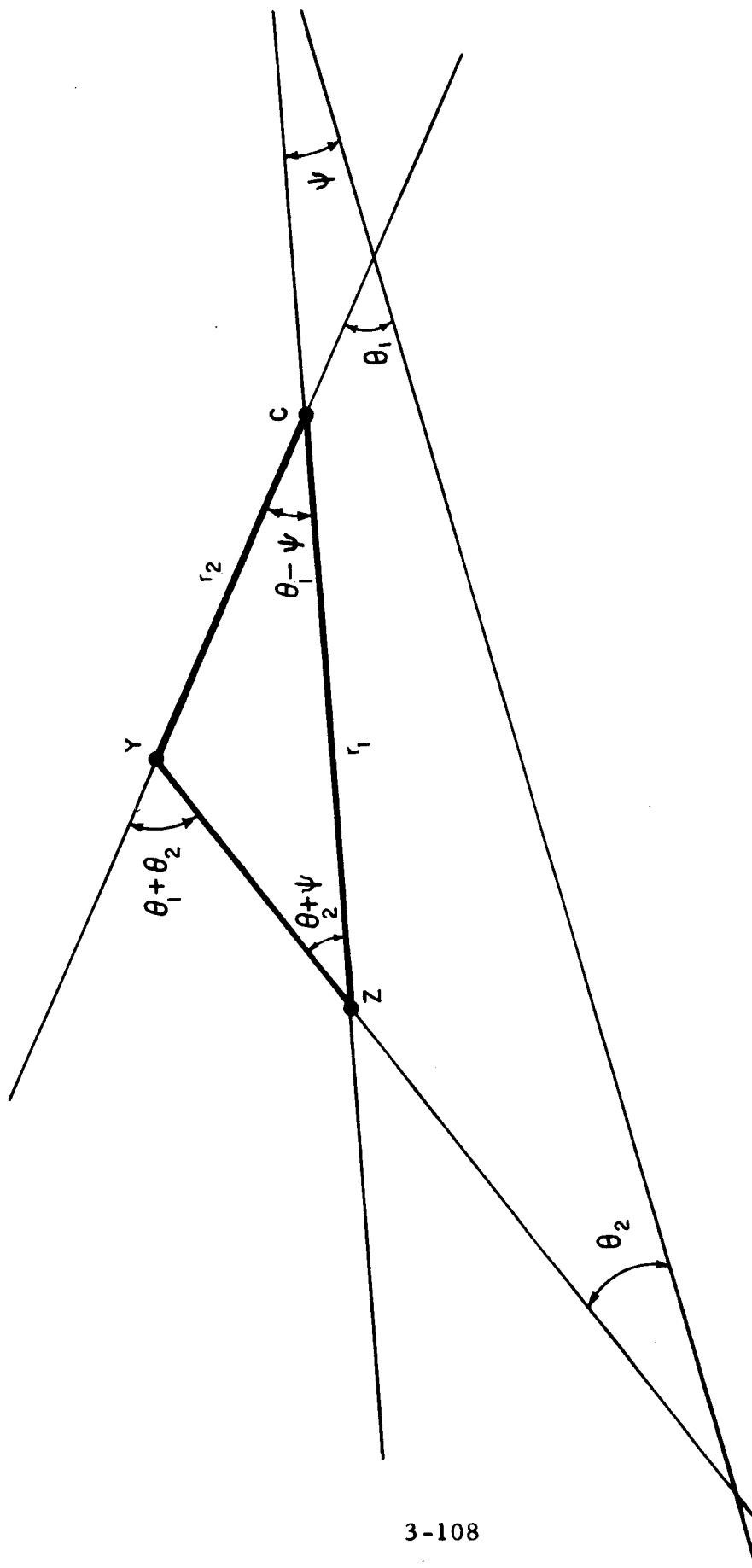


FIG. 3-36

MEASUREMENT GEOMETRY

Error Analysis

The principal source of error, which will define the accuracy of the instrument, is the finite sharpness, regularity, repeatability, etc., of the iris edge of the pupil.

Some close-up photographs of a human eye were taken, in order that some estimate could be made concerning the sharpness, circularity, etc., of the pupil/iris boundary. (Figs. 3-37, 3-38, 3-39). The coordinates of approximately 20 points on the pupil iris boundary were measured with a shadowgraph and then plotted on graph paper. Various circles were drawn in an attempt to fit these points. A circle was located on which these 20 points were randomly distributed on either side of the circumference. (Illustrated in Fig. 3-40). This corresponds to the best fitting circle. The uncertainty in the position of the center of a set of 20 points is approximately equal to $\frac{1}{\sqrt{20}}$ x ORMS deviation of the 20 points from the circumference. This was found to be, approximately, equivalent to an angular error of 10 minutes of arc.

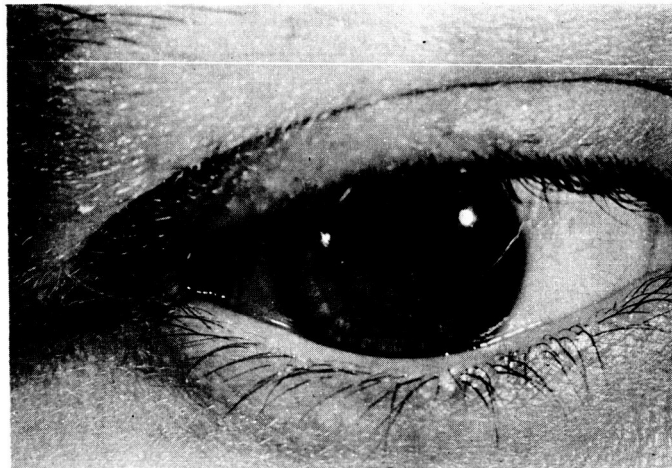
This result demonstrates that the pupil circle is defined to an accuracy that is sufficient to meet the requirements of the instrument. (In an actual instrument a much larger number than 20 points would be sampled and the accuracy would be correspondingly better).

The most probable source of error of the instrument will be due to random displacements of the pupil center following expansion and contraction of the pupil. This effect has not been measured but it is not expected to lead to errors substantially greater than that due to the unevenness and finite sharpness of the pupil boundary.

The accuracy with which the pupil may be tracked depends on how well the pupil is defined (i. e. the considerations discussed above) and also on factors such as the amount of light available, the speed of response of the instrument, the quantum efficiency of the photodetector, and the sharpness of focus that can be maintained by the pupil tracker. This latter, is related to the depth of focus of the pupil tracker optics and to the accuracy with which the instrument can be located relative to the eye.

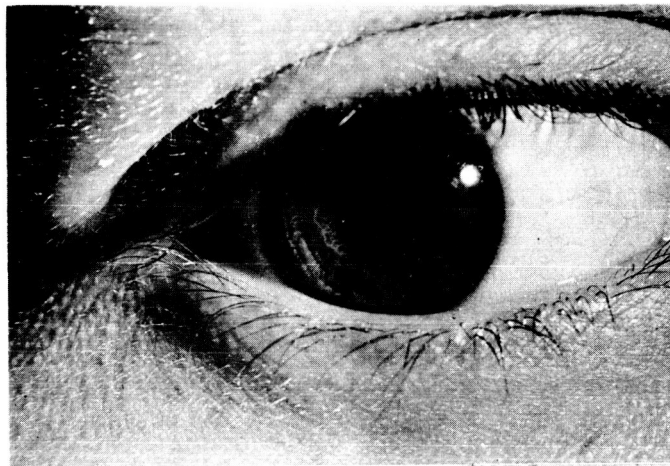
PUPIL / IRIS BOUNDARY DEFINITION

FIG. 3-37



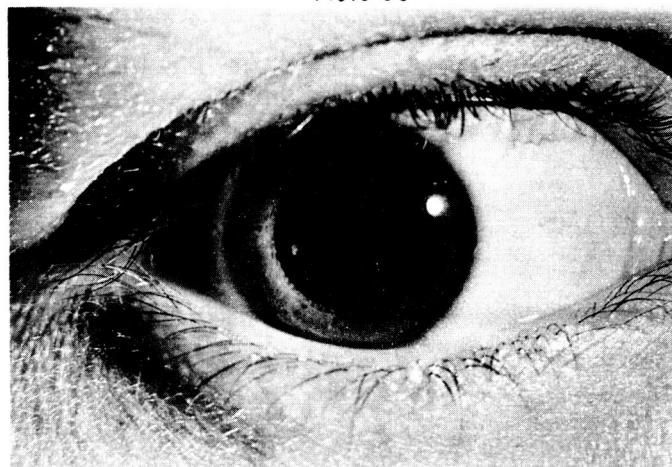
PUPIL ACCOMMODATED TO BRIGHT LIGHT

FIG. 3-38



PUPIL ACCOMMODATED TO MODERATE LIGHT

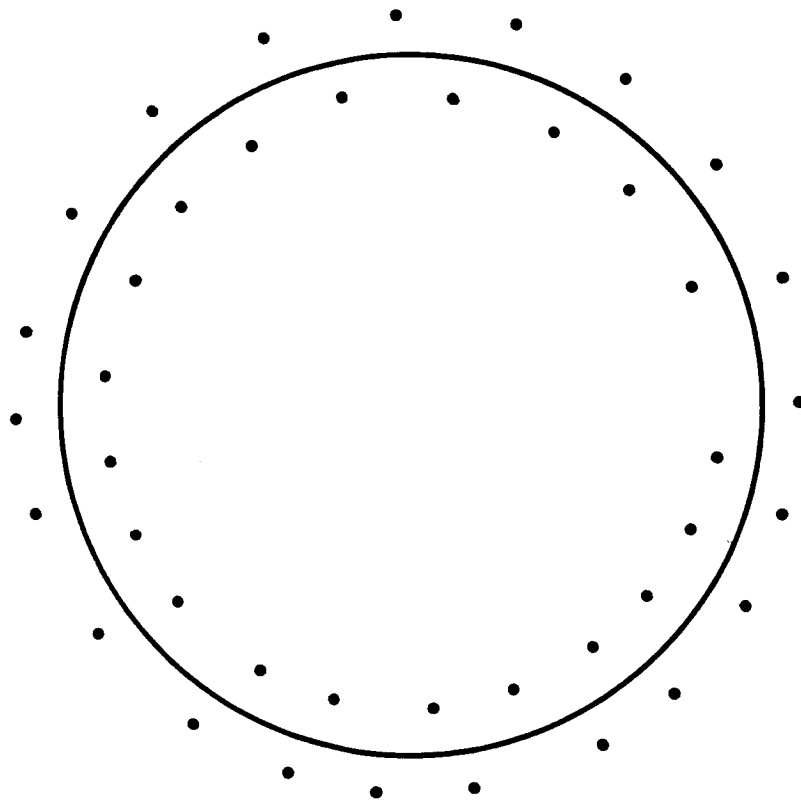
FIG. 3-39



PUPIL ACCOMMODATED TO DIM LIGHT

FIG. 3-40

**BEST FIT CIRCLE TO A SET OF POINTS
(ILLUSTRATIVE ONLY—NOT TO SCALE)**



REFERENCES - SECTION 3

- 3-1 Healer, J., Arms, A., Merchant, J., and Britt, E., A Summary Report on a Review of Biological Mechanisms for Application to Instrument Design. ARA 1025, Allied Research Associates, Inc., January, (1962)
- 3-2 Davson, H., The Eye. Academic Press, New York, (1955)
- 3-3 Le Grand, Y., Light Colour and Vision. John Wiley and Sons, New York, (1957)
- 3-4 Fry, G.A., Handbook of Physiology, Section I, Vol. I, Chap. XXVII, American Physiological Soc., Washington, D.C., (1959)
- 3-5 Ludvigh, E., and McCarthy, E.F., "Adsorption of Visible Light by the Refractive Media of the Human Eye," Arch. Opthal., New York, 20, 37, July (1938)
- 3-6 Flamant, F., "Étude de la répartition de lumière dans l'image retinienne d'une fente" Rev. Opt. (theor. instrum.) 34, 433
- 3-7 Rosenblith, W.A., Sensory Communication. Chap. 21, M.I.T. Press, (1961)
- 3-8 Westheimer, G. "Optical and Motor Factors in the Formation of the Retinal Image," J. Opt. Soc. Amer., p. 86, January, (1936)
- 3-9 Nachmias, J., "Two-Dimensional Motion of the Retinal Image During Monocular Fixation," J. Opt. Soc. Amer., 49, 901, September, (1959)
- 3-10 Riggs, L.A., Armington, J.C., and Ratliff, F. "Motion of the Retinal Image During Fixation," J. Opt. Soc. Amer. 44, 315, April, (1954)
- 3-11 Westheimer, G. "Eye Movement Responses to a Horizontally Moving Visual Stimulus," A.M.A. Arch. Opthal. 52, 932, July, (1954)
- 3-12 Stark, L. and Young, L.R., "Predictive Control of Eye Tracking Movements," IRE Transactions on Human Factors in Electronics, p. 52, September, (1962)
- 3-13 Young, L. R., "A Sampled Data Model for Eye Tracking Movements," Sc. D. Dissertation, M.I.T., (1962)

4. Chemoreception

In the first summary report (see Ref. 4-51) a survey was made of the various theories and mechanisms pertaining to chemoreceptor response with the initial emphasis on gustation. This year further analysis was made of these theories in addition to new ones, with the emphasis placed on olfaction. Potential applications for instrument design are also discussed.

4.1 Mechanisms of Olfaction

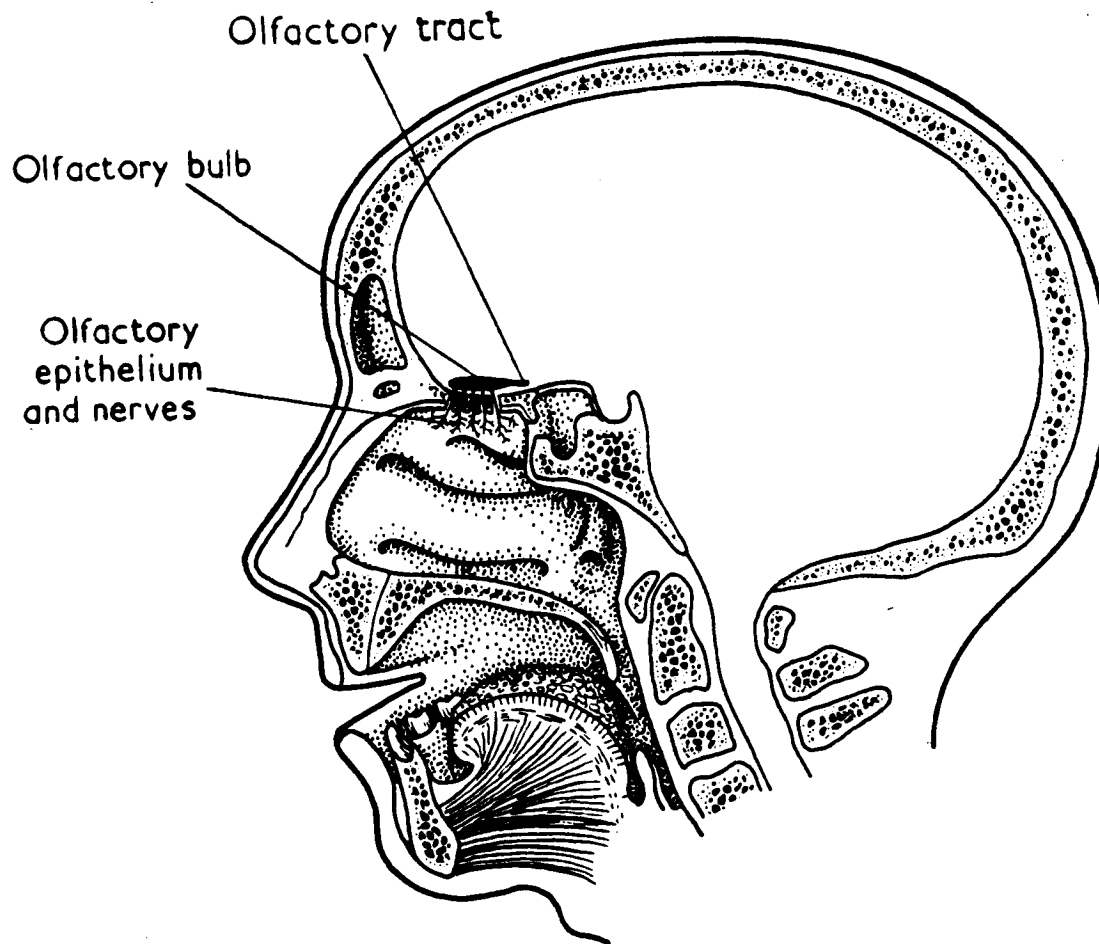
In considering the human olfactory process, it is necessary to look at the entire path that inhaled air follows in its passage through the respiratory region, since the particular section which is highly sensitive to odors constitutes a very small area, and is really quite remote from the incoming air stream. (See Fig. 4-1) It is possible that the whole nasal region influences and possibly participates in the perception of odor. Some quantitative experiments performed by Moncrieff in 1955 (Ref. 4-1) and Stuiver in 1958 (Ref. 4-2) have shown that the mucous membrane or epithelium in this entire region has strong adsorption tendencies towards odors. The number of odorous molecules actually passing over the olfactory sensory area are much fewer than those entering the nostril. Stuiver has estimated that if the threshold for perception of an odor such as secondary butyl mercaptan is approximately 10^9 molecules per second entering the nostril, only 2% or 2×10^7 are adsorbed effectively on perhaps 4×10^7 olfactory sense cells. This figure is a rough estimate, but it does serve to indicate that the number of odorous molecules necessary to excite an individual sense cell may be fantastically small, perhaps as low as 1 to 8 molecules per cell. The latter also implies that all the odorous molecules which reach the olfactory area actually contact the sensory receptors themselves. If the latter is the process that occurs, at least four different mechanisms are possible.

1. Simple contact of odorant and receptor.
2. Chemical reactions between odorant and receptor.
3. Chemical or physical adsorption of odorous molecules on the receptor surface.
4. Dissolution of odorous molecules in the membrane or tissues of the receptor.

Theory No. 1 implies that the odorous molecules simply hit the receptor surface and bounce off instantaneously. In this case pressure or

FIG. 4-1

Section of head to show nasal cavity.



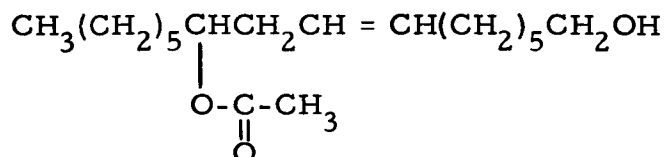
vibratory effects would have to be given primary consideration. However, if this is the case, olfactory sensations from the passage of pure air would have to be felt and this does not occur under normal circumstances. Mechanisms No. 2, 3, and 4 probably play separate roles which taken together may initiate the olfactory process. A strong likelihood exists that olfactory cells retain a highly active surface layer of protein-like substances (enzymes or pigments) which due to their great complexity either react themselves with odorous molecules, or influence a chemical reaction such as oxidation, reduction, hydrolysis, etc. between the odorant and the cellular or cilia membrane. It is also possible that the surfaces of olfactory cells may be so biologically active that no separate enzyme is needed to initiate chemical reactions. However, in considering the morphology and chemical composition of various vertebrate olfactory epithelia which are usually found to be yellowish in color (in contrast to the reddish color which is seen in the liquid sheath or mucous membrane that constitutes the remainder of the respiratory region) it is found that protein-bound carotenoids, free carotenoids, and free vitamin A abound in these layers. Carotenoids belong to four general types - unsaturated hydrocarbons, ketonic or hydroxylic derivatives, acids, and esters. They usually occur in complex mixtures and are quite active chemically. Briggs and Duncan (Ref. 4-4) have postulated an analogy with the protein-bound haplocarotenoids of the retina, i. e. the rhodopsins (where the action of light appears to induce cis-trans isomerization prior to the initiation of the nerve impulse (Ref. 4-5)). They assume that the extensive conjugated system of olfactory carotenoids are the receptors of energy from olfactant molecules, thereby causing similar cis-trans molecular changes to occur. However, the exact chemical-mechanism (which may cause depolarization of the receptor cell membrane) involved in this process was not discussed. The carotenoid-protein theory of olfaction has been criticized by Moulton (Ref. 4-6) but the possibility of its existence cannot be ruled out. The question of chemical reactions taking place will be explored again later in this section.

4. 1. 1 Monolayer Theories

Theory No. 3 mentioned above undoubtedly occurs to some degree and may be tied in with the enzyme theory, since adsorption must take place prior to the initiation of a chemical reaction. In any case, it is assumed

that the receptor surface has an attractive force for certain types of gas molecules and is able to hold them momentarily until either the vibrations of the surface atoms impart a sufficient energy to the adatoms in order for them to evaporate back into the gas stream, or they dissolve in the membrane. The longer the gas molecules remain on the surface the greater will be the concentration of the gas that is adsorbed or being built up, but this may not be essential. It would be desirable to know the value of the heat of adsorption of various odors on the receptor surfaces. For example a ΔH of 10 kcal per mole corresponds to approximately 10^{-6} seconds of contact with the surface. This means that in the case of mercaptan odor perception previously discussed, where perhaps 10^7 molecules are in contact with the surface every second, the actual equilibrium value for the number of molecules adsorbed at any given instant would only be 10 molecules. This, of course, is such a low figure that it seems much more likely that an adatom on contact with the receptor surface has a stay of at least several seconds. The latter would imply a heat of adsorption close to 20 kcal per mole. Thus, we are actually approaching the region of chemisorption where the mechanism of adsorption would be different than that in the physical adsorption region. In other words, it could be quite possible that electron transfer of a particular nature (similar to an actual chemical compound) takes place between odorous molecules and the surface. If only physical adsorption were considered, this type of electron transfer would probably not occur and only polarization effects would have to be considered. A heat of adsorption that is as large as 20 kcal/mole could also indicate that a monolayer of odorous molecules is continually being built up. Consider, for example, the unlikely case that the olfactory epithelium consists completely of olfactory sense receptors at the surface, which is an active area of 2.5 square centimeters (the area of the olfactory region has been measured). It can then be seen that it would take of the order of 10^{+7} - 10^{+8} seconds for a complete monolayer to be adsorbed (10^{14} - 10^{15} adsorption sites/cm²) if it is assumed that the molecules remained on the surface for an infinite period of time. If the lifetime of adsorbed odor molecules is extremely short (<1 sec.), equilibrium conditions would be realized instantaneously and a monolayer would never have the opportunity to be built up. It should also be emphasized that if the build up of a monolayer takes place

reactive than less odorous or non-odorous molecules. Various osmophoric groups probably contribute substantially to the odor producing characteristics of various organic vapors. For example, a substance studied in great detail in terms of its olfactory characteristics is the sex attractant produced by the female Gypsy Moth. Males of the specie are extremely sensitive to its presence. There is evidence that they can detect a single molecule. The attractant has the following chemical structure:



Unsaturation is present, a methoxy and carbonyl group are close together, and the molecule has a very long chain. These factors taken together (the influence of the hydroxyl group is probably more speculative) may be quite favorable in the odor producing characteristics of the molecule. Obviously, the male moth has chemoreceptors which are highly sensitive and selective for detecting this complex molecule.

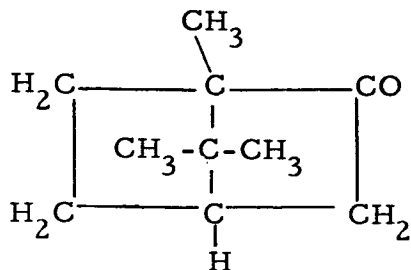

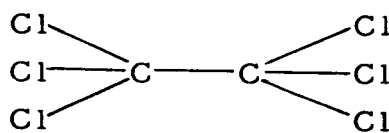

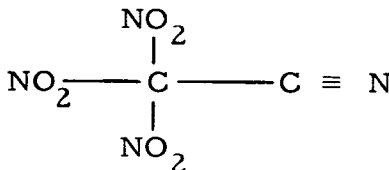

4. 1. 2 "Lock and Key" Concepts

The concept of specific odor receptors and adsorption phenomena in general have led many theorists (Ref.4-8 through 4-17) to consider a "lock and key" approach as being of prime importance in the olfactory mechanism. This theory considers that a volatile complex organic substance must contain molecules of a prescribed three dimensional shape and size which will fit on certain available molecular sites (also possessing definite shapes and dimensions) in the olfactory receptor membrane. Analogies to this "lock and key" concept may be found in various dying processes whereby some dye-stuffs of certain molecular dimensions will dye a fiber, but if their chain is lengthened then they will not do so, although all the necessary osmophoric groups are still present. It is thought that this is a case of the dye-molecule fitting the available sites.

It is a well known phenomenon of olfaction that chemicals with completely unrelated empirical or structural formula exhibit very similar odors. For example, if the following three substances are considered - such as d-camphor, hexachloroethane and trinitroacetonitrile, it may be

observed that while all three have a camphoraceous odor they appear to have very little in common in regards to their empirical and structural formula

TABLE I

Compound	Empirical Formula	Structural Formula	Stereochemical Formula
d-camphor	$C_{10}H_{16}O$		
hexachloroethane	C_2Cl_6		
trinitroacetoneitrile	$C_2N_4O_6$		

However, the stereochemical models (TABLE I) of these molecules show that they have quite similar spherical-type shapes (Ref. 4-14). In addition, each has a size close to 7\AA in diameter. Amoore has looked at more than 100 chemicals with a camphoraceous odor, and found that almost all of them conform to this shape and size within quite narrow limits. In fact a total of more than 600 odorants which can be broken down to the seven primary odors of ethereal, camphoraceous, musky, floral, pepperminty, pungent and putrid have been investigated, and few exceptions to the theory were reported (Ref. 4-13). Each primary odor is thought to have its own stereochemical configuration which, at least on the human olfactory epithelium will fit receptor sites possessing compatible three dimensional structures. More than one receptor site may be distributed over each olfactory cell with the result that

one major difference between cells would be the existence of different ratios of the seven primary receptor sites. In this case the variability that is seen in the response to numerous odors would then depend on the relative concentration of the different sites and on the occupation of a threshold number of sites of more than one type.

4.1.3 Morphology of Respiratory and Olfactory Region

In the first summary report (Ref. 4-51) it was shown that the olfactory receptor cell is a bipolar neuron with a peripheral extension from which a number of hairlike filaments protrude into the mucous covering the epithelium. The afferent nerve fiber emerges from the proximal pole of the cell and runs together with fibers from adjacent cells to the brain. One of the basic problems in the study of the function of the olfactory sense organ concerns the question of the exact part of the primary neuron that serves as the sensory element. Various parts of the cell, such as the olfactory rod, the vesicle or the hairs have been suggested to represent the chemosensitive portion of the cell. Experimental evidence appears to support the view that the primary reaction between the stimulating agent and the neuron takes place in the membrane of the hairs. Thus, this entire surface network might be regarded as both the sensory membrane and chemoelectrical transducer of the olfactory apparatus. However, it should be emphasized that the time of response of the olfactory epithelium (membrane potentials) to excitation appears in many cases to be relatively slow compared to other sensory organs. This could be due to the fact that the exciting odorous molecules have to pass through a layer of mucous before they reach the receptors. Excitations might then be assumed to take place gradually with the time course being to a large extent a function of the number of molecules which reach the receptor per unit time. Another factor which may influence the temporal course may be the length of time that it takes for the odorous molecules to become inactivated after being adsorbed. In addition, the aqueous and lipoidal solubility characteristics of odorants may also affect excitation processes, since the rate of passage or the strength of "entrapment" in the mucoid layers and the subsequent reaction or passage through the receptor membranes may be quite dependent on these solubility properties. This ties in with olfactory mechanism No. 4, to which references have been previously made. A survey of the solubility properties of odorous molecules indicate that they must be soluble in

water or other cell constituents. The general trend seems to indicate that olfactory sensations decrease as water solubility increases and lipid solubility decreases. It has been proposed that water soluble substances stimulate the olfactory receptors via a water phase, while many larger molecules only act through a lipid phase. Witheridge (Ref. 4-18) has suggested that as the odorant dissolves in the nasal epithelium the bonding angles which unite the atoms in the odorous molecule are modified. He believes that it is this modification which provides the stimulus for odor perception. In any case, concentration equilibrium conditions are probably quickly established between the media of the inspired air, the mucous, and the receptor sites. In conjunction with these equilibria an investigation of the kinetics of stimulus transport to the olfactory receptors could present many interesting possibilities for further studies.

The preceding discussions seem to imply that various opinions are in accord as to the structure of the olfactory epithelium surface. However, this is far from the actual situation as will be shown. While the majority of present day investigators consider that the rods of the olfactory cells protrude out to the epithelium-air interface, there are some, such as Mateson (Ref. 4-19) who after a careful investigation came to the conclusion that there is no evidence that this phenomenon occurs (at least in humans) and therefore the olfactory elements themselves do not actually contact the gaseous environment to which they are exposed. (The olfactory area in vertebrates is seen to contain three types of cells of which the olfactory cells constitute 70% of the total). In any case, whether olfactory receptors are located at the air cavity interface itself or at a certain distance away from this interface, there does not appear to be much doubt that we are in essence dealing with some type of liquid sheath at the surface. It is also quite probable that this liquid sheath is continuously being replenished with fresh layers.

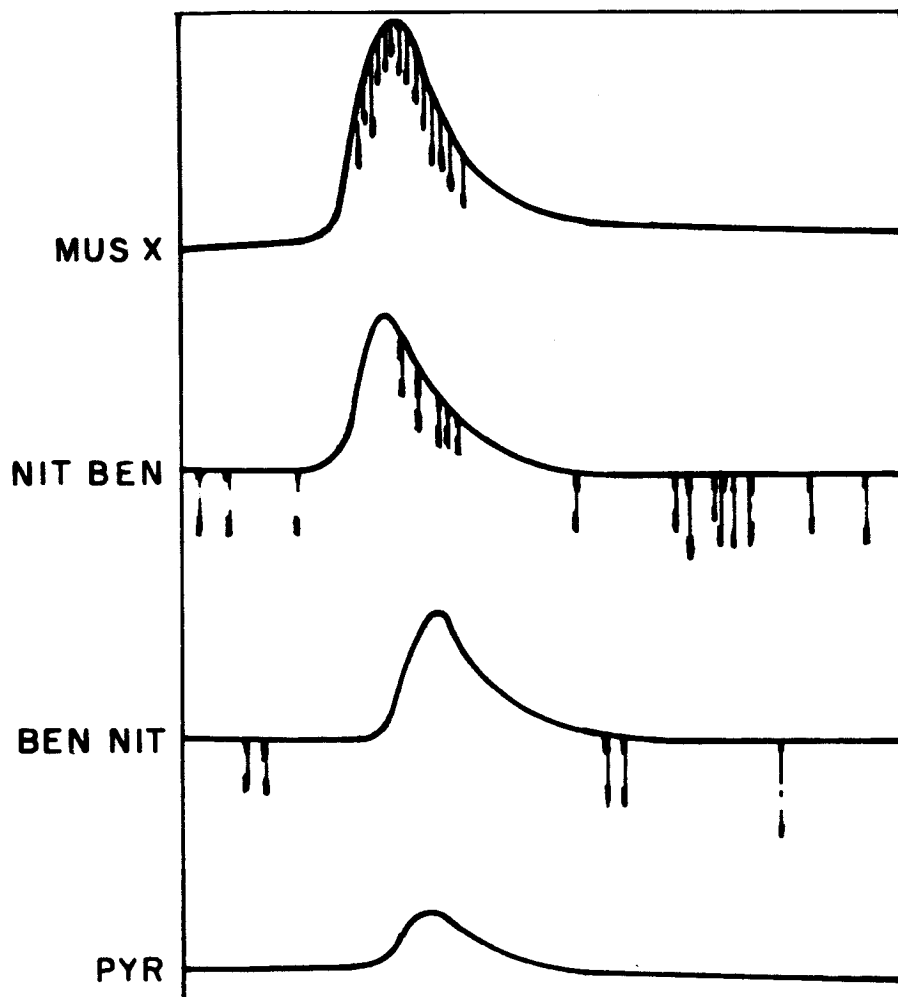
4.1.4 Electrophysiological Considerations

Brief mention has previously been made concerning various electrophysiological experiments on the electrical activity of olfactory epithelia. Since even the smallest microelectrodes that were utilized in the experiments were too large to record activity in one individual olfactory receptor (even if the actual receptor site were known), the basis of the experiments was the recording of slow and fast "action potential" spikes produced by the

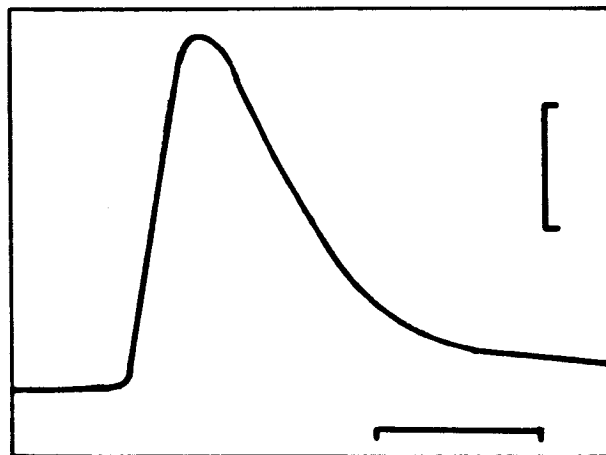
passage of odors over the area. The experimental difficulties involved in this experiment have been substantial, and difficulties in interpreting the pattern data are frequently encountered. However, recently with the development of newer techniques and smaller microelectrodes, the odor-specific properties of small groups of cells can be differentiated. A recent paper by Gesteland (Ref. 4-15) has described a method of recording the action potentials of olfactory receptors by using low-impedance platinum-black plated, metal filled microelectrodes which are coupled capacitively to a cathode follower. Small puffs of odorized air from 1 ml syringes were blown directly into the mucosa of a frog and clearcut responses were obtained from different areas. Many different odors as well as stimulus strengths were utilized in the experiments, with some of the resulting oscillograms shown in Figure 4-2.

The results of these experiments together with others appear to indicate that either odor discrimination depends to a large extent upon the existence of variations in the composition of the mucoid layers in different types of olfactory receptors exist with different sensitivities to basic odors which, individually, possess particular molecular configurations. An introduction to the "lock and key" concept of odorant-receptor interaction, which is quite popular with the majority of physiologists, has been previously made. Adrian (Ref. 4-20 through 4-22) has considered that both spatial and temporal patterns are of equal importance in contributing to an overall "signature" of various smell patterns. He also found that the time course of discharge as recorded will vary in different groups of cells. The importance of threshold concentrations of odorant molecules should also be of prime importance in olfaction since it was previously thought that the entire respiratory area can serve as an effective adsorbant for odors. Adrian's investigations on the human olfactory epithelium showed that with substances very soluble in water the threshold for odor perception was lower in the anterior part of the olfactory organ, whereas with substances mainly soluble in lipoids the threshold was lower in the more remote region. One possible way in which different areas on the olfactory surface may differ from each other, might arise from differences in the packing of receptors in various regions. This might take the form of some receptor cells having their nuclei nearer the terminal end than others (closer to the epithelium-air interface) which might cause a difference in the time of response to odorant stimuli to occur.

FIG. 4-2
RESPONSE TO MUSK XYLENE
NITROBENZENE, BENZONITRILE
AND PYRIDINE



RESPONSE TO BUTANOL VAPOR



This section has explored specific parameters that contribute to olfaction such as spatial patterns, temporal patterns and specific receptors. The whole range of olfactory discrimination is further enhanced by the role of the central nervous system which can integrate all the incoming signals and produce the appropriate response.

4.1.5 Radiation and Vibrational Theories

In the last summary report (Ref. 4-51) brief mention was made concerning various theories relating electromagnetic radiation to olfaction. The infrared and Raman spectra of many odorous molecules have frequently been studied on the assumption that intramolecular vibrations cause resonance or interference with the vibrations of the olfactory surface. Similar odors are frequently found to have similar radiation absorption bands. Recent work concerning this "vibrational theory" has been published (Refs. 4-23 - 4-29). However, the popular hypothesis is that odors emit low frequency vibrations. Characteristic molecular frequencies should, therefore, be present in substances known to be bio-active. A search has been made for low frequency vibrational modes in these molecules using models and published structures (Ref. 4-24). The probable interactions between atoms or groups and the possible effects of hydrogen bonding were taken into account. For example, in considering the Gypsy Moth sex attractant previously mentioned, the -CO-CH_3 group is thought to be pivoted in such a way that the -CO- can oscillate up and down with reference to the locally rigid reference frame provided by the double bond. Because of the bonding of the -CO- to the -OH , there is likely to be a considerable electronic displacement accompanying the oscillations so that a relatively strong band may be expected in the far infra-red spectrum of this substance. Also in comparing the "cis" and "trans" form of this complex molecule, it appears that while the same type of hydrogen bonding of the hydroxyl and carbonyl groups may take place, the orientation of the olefinic linkage will be different enough to make the elastic forces (and hence the frequencies) different in both isomers.

In summary, Wright postulates that a number of different types of receptors would be necessary to account for the very high informational capacity of the olfactory apparatus in vertebrates; each type being "tuned" to a narrow frequency band and generating a nerve impulse when

approached by odorous molecules vibrating with the correct matching frequency. Although, in vertebrates, sensitivity to a substantial number of distinct vibrational frequencies may be the usual case in olfaction, exact correlations of odors with vibration is an extremely difficult job. However, in insects, while the basic mechanism involved may be quite similar, it is probable that it is very much simplified since the insect brain does not demand an olfactory apparatus of high informational capacity. All it may need is a simple "yes" or "no" to tell it that a certain chemical scent is present or not, and the appropriate behavioral response would then follow automatically. For that purpose Wright believes that a sensory apparatus tuned to a fairly narrow band of frequencies would be perfectly adequate. Obviously, much careful experimental work remains to be done before any definite conclusions can be made concerning the plausibility of a vibration theory. In any case, it still appears most likely that actual contact between the odorant molecules and the chemoreceptors surface must take place prior to the initiation of the nerve impulse. One recent theory (Ref. 4-30) suggests that odor results from the adsorption by odorous substances of infrared rays emitted by the olfactory area with a subsequent transformation into nervous impulses being initiated by a change in thermal equilibrium. Different odors were considered to be the result of different patterns of infrared adsorption. However, experiments have recently been performed which seem to cast doubt upon this theory (Ref. 4-31). The olfactory epithelium was coated with a thin plastic membrane permeable to infrared radiation, and various odors were passed over this region with no apparent change in electrical activity being recorded. Perhaps the Raman spectra has more of an effect, especially with substances such as carotenoids which, as previously seen, may play a very important role in the olfactory process. However, if the assumption is made that some sort of adsorption process must initially take place in the nasal region the possibility then occurs that other parameters (in addition to characteristic vibrational frequencies) are important in the initiation of olfaction, such as polarizabilities, dipole and dispersion type interactions, bond moments, electron transfer, etc. These various properties can contribute to differences in adsorption specificity for various epithelia-odorant interactions, especially if the adsorbed molecules possess particular osmophoric groups. Dravnieks (Ref. 4-32) feels that the relative rapidity of appearance and disappearance of the

olfactory sensation indicates that the energies of adsorption must be in the range of 5-15 Kcal/mole (compare with estimation of 20 Kcal previously discussed). Typical adsorption isotherms might then show coverages less than 0.0001% of a monolayer. This might indicate that as few as 10^5 - 10^7 molecules of certain odorants could be detected. Of course, the surface potential change on the receptor surface would under normal circumstances be quite small; however, Dravnieks postulates that an amplification of such a change is possible on a ferroelectric substrate. Below the Curie point, before the onset of spontaneous polarization, a mechanical strain apparently exists in ferroelectric lattices. If a sufficient number of selective odorous molecules adsorb on receptor sites, the polarizing field of the permanent or induced dipoles of the osmophoric groups could serve to nucleate the lattice shear, thus triggering the ferroelectric transformation throughout the entire domain. This would then result in a spontaneous polarization, with a so-called Barkhausen pulse (Ref. 4-33) possibly initiating the olfactory nerve impulse. In other words, the orientation nucleated by the adsorbed molecules may spread through the lattice and cause a separation of charges. The olfactory receptor may be restored to the nonferroelectric lattice form periodically by a metabolic process followed by the transformation back to its "strained" condition prior to the initiation of a new pulse. The existence of ferroelectric substances in living organisms has been shown by Athestaedt (Ref. 4-34), among others. Some of these substances such as guanidine, glycine sulphates, and thiourea can be considered to be in the class of organic semiconductors. Dravnieks speculates that a strong possibility exists that this overall theory may occur in proteinic structures which of course comprise some of the major chemical components of receptors in general.

Analyzing the various dipole theories of olfaction, as well as those theories which include vibrational movements, it should now be realized that these theories can be quite compatible with the lock and key concepts previously considered. The stereochemical configurations of the odorant molecule and receptor site can play a major role in determining the closest distance of approach, thus influencing the effective moment of the functional group as well as possibly supplying the necessary additional increments in the energy of adsorption to shift its value onto a pulse-stimulating band.

The following Table taken from Moncrief's The Chemical Senses presents the salient features of many theories pertaining to odor through 1949.

TABLE II

Author	Date	General Class	Salient Features
Ogle	1870	Vibrational	Vibrations affected nasal pigment. which gave out heat which excited the olfactory cells.
Woker	1906	Chemical	Unsaturation main cause of odour, but not essential if substance very volatile.
Fabre	1911	Vibrational	Limited to insects. Not known by man. Human olfaction due to material particles.
Marchand	1915	Chemical	Unsaturation (including $>C=O$). Two points of unsaturation reduces odour.
Henning	1916	Chemical	Osmophore groups are important, but their relative position determines the type of odour.
Heyninx	1917	Vibrational	Vibrations causing absorption in the ultra-violet band also caused odour.
Backman	1917	Chemical	Water solubility and lipid solubility essential.
Teudt	1919	Vibrational	Electronic vibrations of sensory nerves increase by resonance with similar vibrations of odorants.
Durrans	1920	Chemical	Residual affinity. Addition reaction on the olfactory epithelium.
Heller	1920	Chemical	Direct chemical action on nerve-ending.
Ruzicka	1920	Chemical	Osmophore and osmoceptor.
Tschirch	1921	Chemical	Substance must be soluble in air. Loose compound formed with plasma of the olfactory cell.
Zwaardemaker	1922	Chemical-vibrational	Possess odoriphore, be volatile, lower surface tension, lipid soluble. Odoriphore depends on vibrations in molecule.

TABLE II (Con't)

Ungerer and Stoddard	1922	Vibrational	Intramolecular vibrations within definite frequency range. Unsaturation helpful. Interference and resonance effects.
Delange	1922	Chemical	Unsaturation.
Missenden	1926	Chemical	Intensity depends on number of molecules making contact with nose. Quality depends on nature of reaction between odorous molecules and lipoid tissues.
Nicol	1926	--	Function of sinuses.
Pirrone	1929	Chemical	Two osmophore groups, one determines type of odour, the other the variety.
Niccolini	1933	Chemical	Volatility. Solubility in nasal mucosa. Oxidizability.
Krisch	1934	Vibrational	Insects
Müller	1936	Physical	Odorous substances are dipolar. Irritate the molecular fields of the osmoceptor in nose.
Dyson	1937	Vibrational	Volatility. Lipoid solubility. Raman shift between 1,400 and 3,500.
Beck and Miles	1947	Vibrational	Infra-red radiation from receptors absorbed by odorants.
McCord and Witheridge	1949	Electro-chemical	Change in bonding angle of odorant molecules on solution in mucosa.
Moncrieff	1949	Chemical	Volatility. Molecular configuration complementary to that of receptor sites.

Most of the theories that have been expounded since 1949 are mainly variations of earlier ones. An analysis of these theories have emphasized the following factors: 1) volatility, 2) solubility, 3) reactivity, 4) vibrational frequencies.

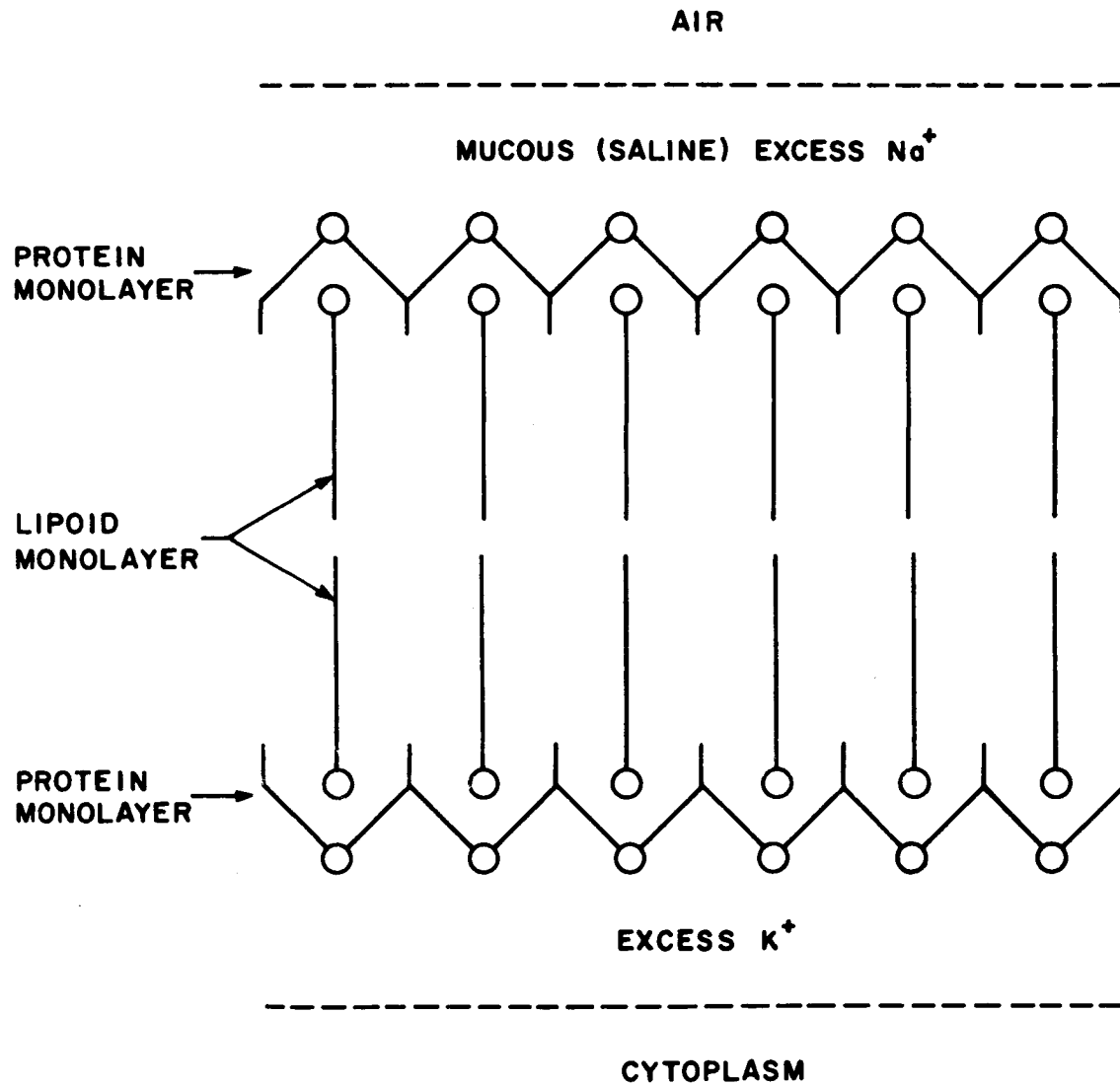
4. 1. 6 Membrane Adsorption and Penetration

One additional approach to the mechanism of olfaction that must be noted was developed during the 1950's by Davies and Taylor (Refs. 4-12, 4-34, 4-35). Their work was based on the earlier experimental observations of Davson and Danielli in 1938 (Ref. 4-36) that certain odorous substances were able to cause a leakage of potassium ions across erythrocyte or red cell membranes which are probably composed of oriented lipid and protein monolayers. This is quite analogous to the behavior of nerve cells in general (detailed discussions pertaining to membrane theory and depolarization effects will be found in a subsequent part of this chapter and elsewhere in the report). Davis and Taylor used the erythrocyte membrane as a model for the membrane of the olfactory nerve cell. They showed that a large number of odorants (in solution, ~

0.005 Molar) are able to act as accelerating agents for haemolysis by saponin (a complex polyhydroxy carboxylic acid) by apparently interacting with some component of the erythrocyte membrane, thus facilitating its disruption. As a model the red cell possesses the experimental advantage that if the cell wall is ruptured, haemoglobin is released with a marked change in the appearance of the solution. In a similar vein, odorant molecules were assumed to be initially adsorbed at the olfactory membrane - mucous interface prior to the initiation of olfaction; using physiological terminology - a depolarization of the membrane potential would momentarily occur as a result of this interfacial reaction and subsequent dislocation at the wall of a receptor site in an olfactory nerve cell (probably at the hairs which form the elongated cilia-type tips at the surface). A diagram indicating the path that an odorant molecule may take from the moment it contacts the air - mucous interface to its subsequent adsorption at the membrane - mucous interface preceding the initiation of a nerve impulse is shown in Figure 4-3.

FIG. 4-3

PENETRATION PATH OF ODOROUS MOLECULE
THROUGH OLFACTORY CELL MEMBRANE



Experimental data (Refs. 4-37 through 4-39) pertaining to the human olfactory threshold, O.T., for many different odors were compared against the overall accelerating power, A, of the odorized solution on erythrocyte cells,

$$\text{where } A = \frac{\frac{1}{T} - \frac{1}{T_o}}{C} \quad (4-1)$$

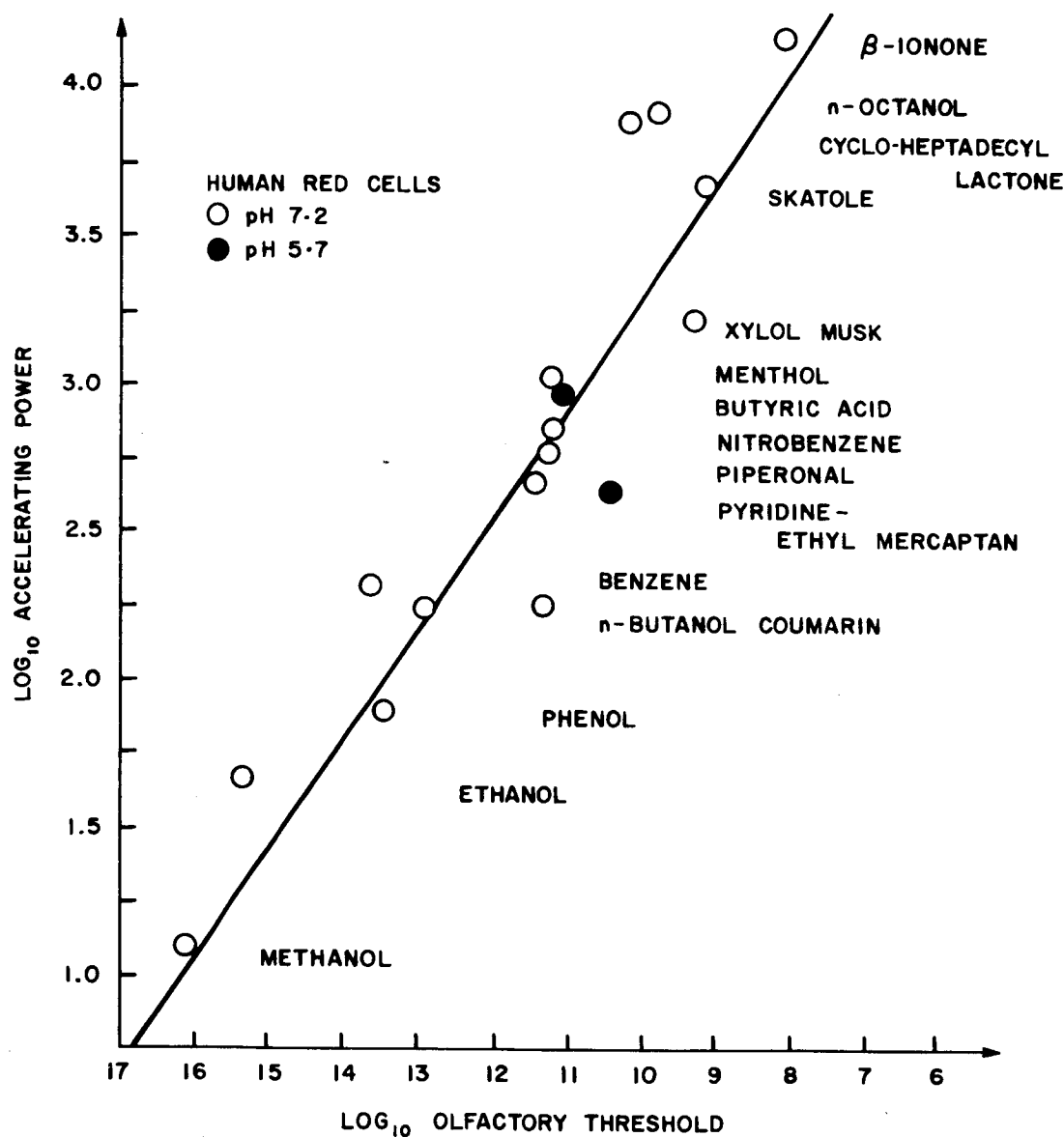
C is the molar concentration of odorant, T and T_o are the time in seconds for the sample to undergo haemolysis, with and without odorant respectively. Figure 4-4 shows a plot of log A vs. log O. T. (in molecules/cc.).

An approximate linear relationship appears to exist between the two parameters. Some of the scatter observed in the plot was attributed to the frequently postulated differences in stereochemical configurations of various odors. Certainly, enormous differences in sensitivity to odorant molecules only slightly different in solubility and chemical properties are frequently observed.

Davies and Taylor considered the application of adsorption theory and molecular morphology to be of prime importance in theoretically deriving a general equation from which the olfactory threshold of any particular odorous compound could be calculated. They divided the olfactory cell surface into the maximum number of sites (each one having a particular area) that could be penetrated or dislocated by complete coverage. In order to "stimulate" the cell, a critical number of odorant molecules must be concentrated on one site on the cellular surface. For the strongest odorants it is thought that only one molecule is necessary to occupy one of these sites, and if a sufficient number of cells had at least one site affected in this manner, olfaction could be initiated. Since weaker odorants apparently have a less "dislocating" effect on the cell membrane, a certain number of odorant molecules, p simultaneously concentrated at one of the sites, may be a necessary condition to evoke a response. The quantity, $\frac{1}{p}$, was then considered a measure of the membrane depolarization ability of the odorant and should have values close to one for very

FIG. 4-4

LOGARITHM OF THE ACCELERATING POWER
OF VARIOUS ODORANTS PLOTTED AGAINST THE
LOGARITHM OF THE OLFACTORY THRESHOLD, AT 25°C.



strong odorants and less than one for less powerful odorants - it has been previously indicated that it may take comparatively few odorous molecules to evoke an olfactory response. Certainly, Davies and Taylor's conclusion as to this possibility is not unreasonable. The olfactory threshold was shown to be dependent on both the energy of adsorption and on p for various odorants. The Langmuir adsorption isotherm for the adsorption of a monolayer of the odorant onto the olfactory membrane was utilized in the theoretical analysis. Figure 4-3 indicates that in order to adsorb on the receptor sites the odorant molecule must pass from the air through the mucous phase. If the process is considered to be reversible at equilibrium, the distribution of molecules can be shown from the following equation:

$$\frac{x}{d} = \frac{cK_{L/A}}{1 + cK_{L/A}} \quad (4-2)$$

where x is the mean number of adsorbed molecules/sq. cm., c is the concentration of the odorant in air, d is the surface thickness ($\sim 10\text{\AA}$) and $K_{L/A}$ is the adsorption coefficient for the odorant in passing from air to the lipid - water interface. Since, in the usual case, c is very small, equation (4-2) reduces to

$$x = cdK_{L/A} \quad (4-3)$$

The Langmuir adsorption isotherm is based on the kinetic theory of gasses and, therefore, factors such as vapor pressure, gas and surface temperature, and molecular weight of the odorant help determine the extent of coverage on the surface. The adsorption constant $K_{L/A}$ can be calculated by making use of the relationship

$$-\log K_{L/A} = \frac{F_{L/A}}{2.3RT} \quad (4-4)$$

where $F_{L/A}$ is the free energy of adsorption, and is equal to

$$F_{L/A} = F_{L/W} - F_{A/W} \quad (4-5)$$

where $F_{L/W}$ is the free energy of adsorption for molecules adsorbing from the mucous (aqueous) phase to the lipid (membrane) - water interface and $F_{A/W}$ is the free energy change for molecules diffusing from air to the mucoid layer. Rideal and Taylor (Ref. 4-12) have experimentally determined values for the free energy of adsorption of various odorants at oil-water interfaces (which can be considered analogous to the olfactory membrane). $F_{A/W}$ can be calculated from the ratio of the solubility of each substance in water to its vapor pressure at 20°C. Free energy values for various odorants as well as experimentally determined olfactory thresholds are given in Table III.

The probability of finding p molecules at a number of sites, N , for a nerve cell can be estimated by the Poisson distribution formula:

$$N = ne^{-\alpha x} \frac{(\alpha x)^p}{p!} \quad (4-6)$$

where the mean number of molecules within each region is x , (previously defined), n is the number of possible sites on nerve surface, and α is the area of a site in cm^2 . If, at any one site, p is greater than the average number of x , it is assumed that olfactory stimulation will occur at the threshold level and therefore $c = \text{O. T.}$ with N equal to unity. With this condition, equations (4-2) and (4-6) lead to the following relation between p and c :

$$\left(ne^{-K_{L/A} c \alpha d} \right) \cdot \left(\frac{(K_{L/A} c \alpha d)^p}{p!} \right) = 1 \quad (4-7)$$

with $c = \text{O. T.}$ and expressed logarithmically the following equation is seen:

$$\log \text{O. T.} + \log K_{L/A} = \frac{1}{p} \log n + \frac{\log p!}{p} - \log \alpha d + \frac{K_{L/A} c \alpha d}{2.3p} \quad (4-8)$$

The last quantity can be neglected since its value is always less than 0.001. If these two extreme cases of a very strong and a very weak odorant are utilized to calculate the necessary constants in equation (4-8) ($\alpha = 64 \text{ \AA}^2$, $n = 44,000$ sites on each cell that can be dislocated, $d = 10 \text{ \AA}$),

TABLE III

Compound	Exptl olfactory threshold molecules/c. c.	$-\Delta G_{O/W}$ cal/mole	$-\Delta G_{A/W}$ cal/mole	$-\Delta G_{O/A}$ cal/mole	Log $K_{O/A}$
Methanol	1.1×10^{16}	1660	5630*	7,290	5.48
Ethanol	2.44×10^{15}	2400	5400*	7,800	5.78
Propanol	5.0×10^{13}	3250	5140*	8,390	6.22
Butanol	8.2×10^{12}	3850	4970	8,820	6.54
Pentanol	6.8×10^{12}	5000	4620	9,620	7.13
Hexanol	8.72×10^{12}	5900	4300	10,200	7.57
Heptanol	9.0×10^{11}	6500	3820	10,320	7.65
Octanol	3.0×10^{10}	7150	4100	11,250	8.34
Decanol	3.6×10^{11}	9360	2760	12,120	8.98
β -Ionone	1.6×10^8	7810	3460	11,270	8.35
Piperonal	2.0×10^{11}	4600	4700	9,300	6.89
Menthol	2.0×10^{11}	6080	2720	8,800	6.52
Skatol	1.8×10^9	6420	5190	11,600	8.60
Xylol musk	2.1×10^9	6180	5350	11,530	8.54
iso Amyl alcohol	6.8×10^{11}	5000	4720	9,720	7.21
iso Butanol	8.2×10^{12}	3350	4840	8,190	6.07
Camphor	5.0×10^{12}	4500	3570	8,070	5.98
Phenol	7.9×10^{12}	4100	6300	10,400	7.70
iso Amyl acetate	1.82×10^{14}	4800	2450	7,350	5.45
Nitrobenzene	2.0×10^{11}	5200	4230	9,430	6.99
Coumarin	2.1×10^{11}	3820	6040	9,860	7.31
Naphthalene	2.6×10^{12}	6300	2460	8,760	6.49
Benzene	4.0×10^{13}	4180	1000	5,180	3.84
Butyric acid	1.4×10^{11}	4300	8320	12,620	9.36
Phridine	3.1×10^{11}	4380	6460	10,840	8.03
cyclo Heptadecyl Lactone	1.78×10^{10}	7200	4600	11,800	8.75
Glycerol	--	2600	5400	8,000	5.93

a final working equation for calculating the olfactory threshold for any odorant molecule becomes:

$$\log O. T. + \log K_{L/A} = - \frac{4.64}{p} + \frac{\log p!}{p} + 21.19 \quad (4-9)$$

Since the p numbers for different odorants probably have a fairly strong relationship to molecular size and shape, correlations can be obtained between the "dislocating power", $\frac{1}{p}$, of each molecular species, and its cross-section area, A . A plot of $\frac{1}{p}$ (p values calculated from Equation (4-9) against the molecular cross sectional area of odorant molecules is shown in Figure 4-5. It is assumed that the odorant molecules are oriented and uncoiled at the olfactory surface. A useful approximation for obtaining the linear plot above is to consider that $p = 1$ for a very strong odorant such as β -ionone and $p = \alpha$ for very small molecules such as water. The line equation that is represented by the plot is

$$pA_o = 47 + 9.4p \quad (4-10)$$

The value of 56\AA^2 for the cross-sectional area of β -ionone is in good agreement with the figure of 64\AA^2 for α , the total area of the receptor site.

Therefore, for calculating olfactory thresholds from Equation (4-9), $K_{L/A}$ may be determined from physical-chemical methods, and values of p may be obtained from Equation (4-10). A requirement for a strong odorant seems to be that a substance must possess a large value of $K_{L/A}$ and a low value of p . A plot of observed olfactory thresholds versus O. T.'s calculated from Equation (4-9) and (4-10) is shown in Figure 4-6. Its linear relationship is quite good and seems to corroborate Davies and Taylor's penetration theory.

One important question to be answered especially with regard to intensity of odors is whether excitation of an olfactory cell by an odorant is an all-or-nothing phenomenon (dependence on the number of cells stimulated) or whether there are different degrees of stimulation of

FIG. 4-5
 PLOT OF $1/p$ AGAINST THE MOLECULAR
 CROSS SECTIONAL AREAS
 OF ODORANT MOLECULES (REF 4-12)

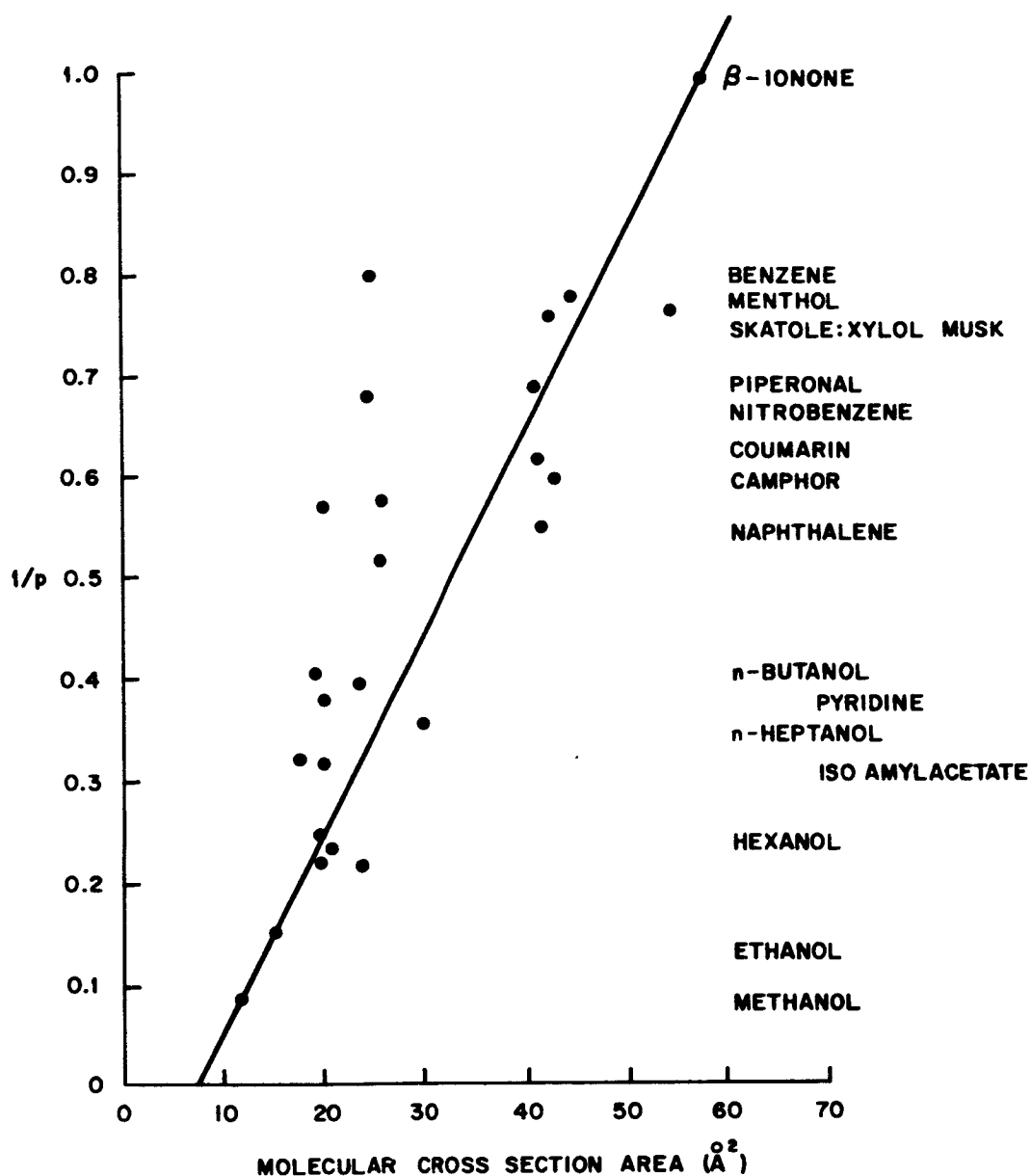
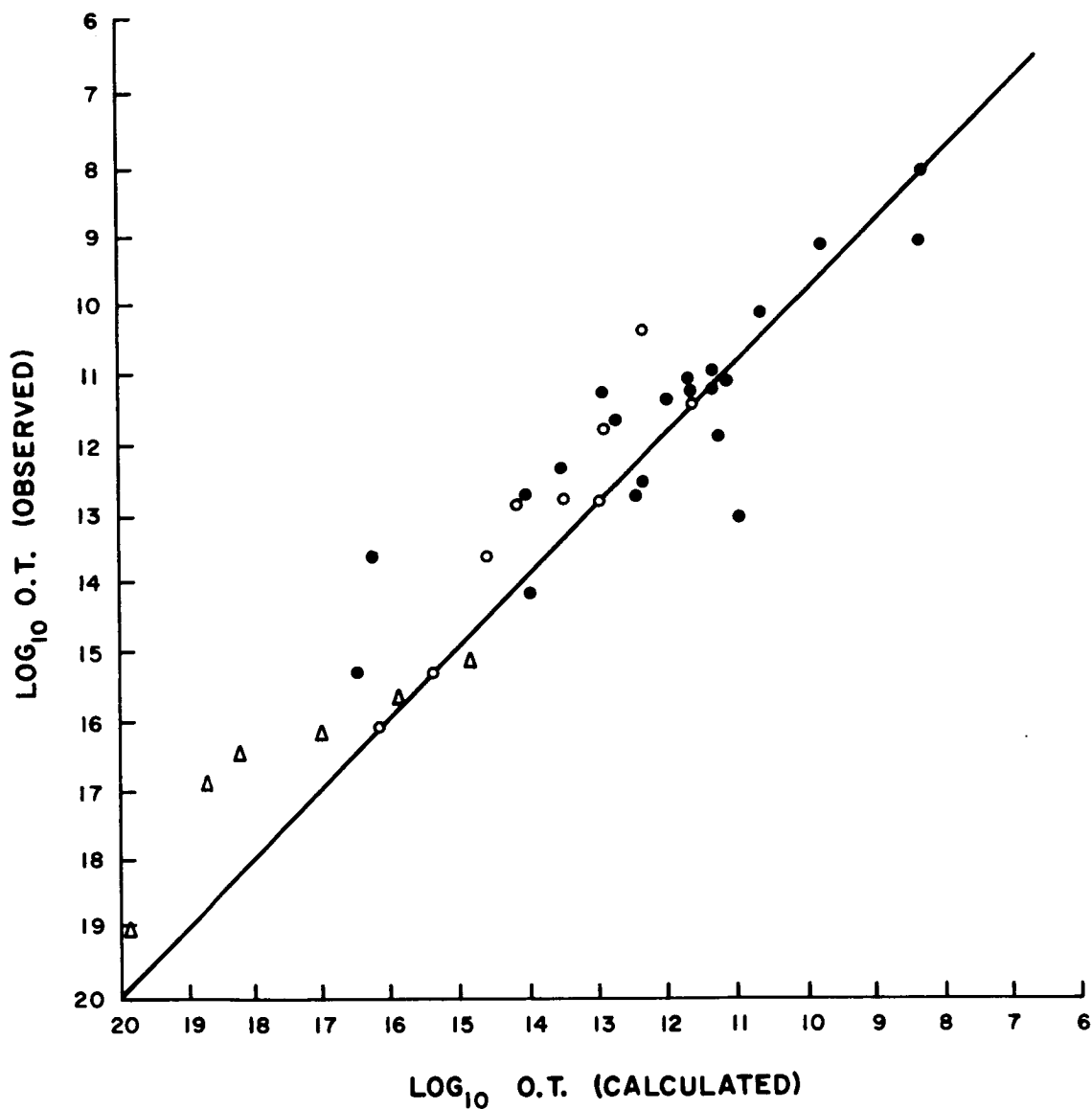


FIG. 4-6

Comparison, on logarithmic scales, of observed olfactory thresholds with those calculated from equations 4(9) and (10). Open circles refer to normal alcohols, triangles to normal hydrocarbons, and closed circles to other organic compounds.



of one cell (dependence upon the number of small areas per cell occupied by p odorant molecules). Of course, other questions that still remain are the importance of spatial and temporal patterns in olfaction, and whether there are actually more than one type of receptor in the olfactory epithelium. One possibility is that there are several types of olfactory receptors but they are specifically sensitive at concentrations only just above the threshold; at concentrations much above this they may react to more odorants.

4. 2 Insect Chemoreception

Investigations of the morphology of various insect chemoreceptors have shown that they apparently consist of well developed sensory cells, but unlike the vertebrates, their receptors (usually on the antennae and hairs) do not appear to be bathed by mucoid layers, and therefore are more readily accessible to the surrounding medium. While the overall sensory system is much more simplified in insects than in vertebrates, the sensitivity threshold and specificity to certain molecules either in air or extremely dilute aqueous solutions are in many cases much greater than that occurring for vertebrates. In the majority of cases observations appear to indicate either a rejection or acceptance of particular chemical stimuli. The high specificity of the receptor sites probably offer a major contribution to this phenomena. For example, selective insect chemoreceptors may accept substance such as carbohydrates while rejecting various kinds of salts. Electrophysiological experiments have revealed the same type of membrane depolarization and action potential that are found in the many vertebrate experiments conducted. However, just as with vertebrates, further analysis of the fine structure and chemical composition of the receptor cell membrane is still necessary. In any case, most of the theories that have been put forth in this chapter (and in the previous report, Ref. 4-51) pertaining to olfaction and gustation appear to apply to insects as well as vertebrates, especially those that consider adsorption pheonmena and the lock and key concept.

4. 3 Sensory Receptors in General

It is necessary to consider the chemical and molecular basis of nerve activity in general in order to obtain a better understanding of the

various chemoreception mechanisms. The biological mechanisms of cells in general were reviewed, and comparisons made between neurons and specialized receptors of all types. By themselves, cells are very sensitive and, depending on their function, are able to detect the most minute change in environmental conditions. It is also quite evident that responses of complex sense organs are determined by the fundamental properties of individual receptor cells, by the influence they exert upon one another, and the control exerted on them by other organs. Thus, the activity of receptor cells is integrated into complex patterns of nervous activity.

One of the parameters that all cells have in common is the plasma membrane which serves as a barrier between the cyto-and expolasm, and is largely responsible for their specificity of function. This membrane is the first component of a cell to react to a change in environment; thus the basis for cell "irritability". The plasma membrane was originally believed to be primarily lipid as a result of cell permeability studies, but largely as a result of studies by surface chemists, the structure as seen in Figure 4-3, is now considered to consist of two oriented lipid monolayers bounded on each end by a complex protein monolayer. Recent studies (Refs. 4-40 through 46) correlating the effects of chemical and physiological agents on membranes and on monolayers have strengthened the monolayer membrane model even further.

This model furnishes not only a valuable tool for analysis of the functional mechanism of various biological receptors but also an important point of departure for application of biosensor principles to instrument design.

4. 3. 1 Mechanism of Drug Action

In many respects the mechanisms in drug action are similar to those described for olfaction, the most widely accepted approaches being that of the lock and key concept and the requirement of equilibrium conditions between the absorbed drug molecule and its concentration in the surround medium. The difference in solubility and thus the permeability of living membranes to various lipid soluble substances helps support the

view that two adjacent monolayers of fatty molecules largely determine the properties of membranes. Experiments concerning the action of certain drugs such as veratrine and its alkaloids and procaine on stearic acid monolayers have shown changes in permeability to potassium and sodium ions (Ref. 4-40), apparently by changes in the surface tension or pressure, F , produced by the molecular penetration, dislocation or reaction with the stearic acid. According to Blank (Ref. 4-47) the entrance of complex organic molecules into a monolayer will require an amount of energy approximating the product of F and the cross-sectional area of the molecule; this can be compared to the relationship between $\frac{1}{p}$ and the molecular cross-sectional values as previously noted. In addition, Shanes (Ref. 4-42) considers that a further dislocation of the membrane (thereby increasing the permeability) occurs by virtue of an increased energy of interaction among the various methylene groups of the membrane molecules which in turn is induced by the increase in surface pressure. One more comment should be made concerning membrane penetration; since many water-soluble molecules are readily able to pass through the membrane, it may be necessary to consider the membrane as some sort of sieve-like structure, with pores or non-lipid passageways, possibly with electrical charges to account for the passage of ions. However, even if this is so, the pores must be assumed to be extremely small and far apart, accounting for perhaps only one thousandth of the total area of the membrane. One alternative to this hypothesis has been the possibility that the function of certain enzymes found on the surface of cells is to convert substances otherwise insoluble in the membrane into soluble derivatives which can permeate through it. In conjunction with this, it should be pointed out that various enzymatic theories of olfaction have also been proposed.

4. 3. 2 Electrochemistry of Membrane Potentials

One of the implications of the preceding discussion is that essentially all living cells have a difference in electrical potential across the cell membrane, and the factors which modify this potential may also modify the activity of the cell. With nerve cells in particular, changes

in the potential are always involved in excitation and propagation of the nerve impulse. Apparently, at rest, under physiological conditions, the inner side of the membrane is negatively charged with respect to the outer side. During excitation the positive charge on the outside momentarily disappears and a wave of electrical negativity synonymous with the nerve impulse spreads along the length of the cell from the region of stimulation. After the passage of the impulse, i. e. the action potential, the membrane is quickly repolarized to its former state. It appears that at least the following conditions are essential to the normal activity of a nerve: (1) metabolism in the cell, giving rise to products which facilitate the diffusibility of ions, especially sodium through the membrane, (2) a differentially permeable membrane through which some ions diffuse inward and outward faster than others, thus giving rise to the steady state "resting potential" or polarization of the membrane (3) reversible changes in permeability of the membrane, permitting momentary depolarization wherein the potential difference across the membrane is momentarily neutralized and then polarized in the opposite direction by the change in diffusibility of ions, (4) a recovery mechanism, following excitation and propagation of the impulse, whereby the resting potential is rapidly restored, accompanied by a reversal of a relatively leaky to a relatively non-leaky state of the membrane; (5) a favorable composition of the exoplasm or solution outside the cell, particularly with regard to sodium, potassium, and calcium ions; (6) the possible relationship between the membrane potential and the rate of secretion of acetylcholine - in axon and synaptic conduction.

Many types of natural and artificial membranes (Ion-exchange phenomena, etc.) show differential permeability to various ions; apparently one important factor that makes the nerve membrane so distinctive is that its permeability is in turn regulated by the potential across the membrane, and it is this mutual influence that is the basis for the signaling process. Certainly, when the voltage difference across the nerve membrane is artificially lowered, the immediate effect is to increase its sodium permeability. The inflow of sodium ions and subsequent outflow of potassium ions is so brief and actually involves such a small number of

that the over-all internal composition of the axon is scarcely affected. Even without replenishment the store of potassium ions inside the axon is sufficient to provide tens of thousands of impulses. At the same time, it must be noted that in the living organism the various enzymatic systems that operate in cellular metabolism serve to keep the "sodium pump" (extrusion of sodium) active with the result that in the resting state the interior sodium ion concentration is held to about 10% of that of the exoplasm (the concentration of potassium ions in the exoplasm is about 30% that of the cytoplasm).

Many of the electrochemical reactions that underlie the resting potential and action potential in nerve activity have been greatly clarified within the past 15 years, although some important questions still remain to be answered. One major problem is the lack of experimental models to simulate the transport mechanisms that occur across living membranes, since biological phenomena are characterized by fantastic complexity and the variables are so numerous. However, utilization of chemical thermodynamic principles, the requirement for equilibria conditions and electrical neutrality, and the resultant electromotive force as applied in the analysis of conventional concentration - type - electrochemical cells, can offer useful models for comparison purposes. It is well known that if two electrodes of the same material are immersed in solutions of their ions at different concentrations, the cell will exhibit a definite electromotive force which depends to a large extent on the ratio of the activities (related to concentration) of the ions in the two solutions. If the different solutions are somehow in contact with each other another potential may develop (called a junction potential) between them. In this case the phenomena of osmotic pressure takes over which causes ions from the more concentrated solution to diffuse into the less concentrated. One factor that determines this junction potential is the difference in the rates of migration of the various ions. The more dilute solution acquires a charge corresponding to that of the faster-moving ions. In many cases equilibrium conditions may be obtained almost instantaneously, thus giving a very brief junction potential. With the interposing of various membranes between the two solutions, equilibria

may be delayed for much longer periods. Artificial membranes have been studied which are either permeable to all ions, permeable only to cations (solution of higher concentration becoming negative to the other solution), or permeable only to anions (solution of higher concentration becoming positive to the other solution). If the membrane is a mosaic of charged patches, some of which are permeable to cations and others permeable to anions, the charge would depend upon the ratio of the two kinds of ions. In any case the greater the mobility of the ions and the greater their penetration of the membrane, the more transitory will be the charge.

If a particular membrane system is set up containing a mixture of ions which include nondiffusible ions on one side and diffusible anions or cations on the other side, a steady state is found to develop which resemble the Donnan Equilibrium for proteins (Ref. 4-48). There is in general an unequal distribution of the diffusible ions between the two sides of the membrane. For example, consider two electrolytes with monovalent ions and a common anion. The ionic concentrations will be given by

$$x^2 = y(y + z) \quad (4-11)$$

where x and y are the activities of the diffusible ions, z is the activity of the nondiffusible ion.

The ratio x/y is termed the Donnan ratio, and can be utilized in a simplified Nernst Equation for calculating the membrane potential:

$$E_M = \frac{\mu_x - \mu_y}{\mu_x + \mu_y} \frac{RT}{nF} \ln x/y \quad (4-12)$$

where μ_x and μ_y are the mobilities of the two diffusible ions,

R is the gas constant,

T is the temperature,

n is the valence,

and F is the Faraday constant.

The differences in electrical potential (as well as osmotic pressure) that exist between the two electrolyte concentrations bounded by the membrane

are seen to be greatly affected by the unequal ionic activities present. As yet it has been quite difficult to apply a rigorous mathematical treatment to nerve membrane potentials during rest, excitation and recovery. However, the Donnan formulation has been applied to predict the movements of Na, K and Cl ions into and out of living cells since these three ions are present in the highest concentration in the cyto - and exoplasm. Hodgkin and Katz (Ref. 4-49) utilizing various simplifying assumptions (such as using the approximation that the potential gradient may be taken as constant throughout the membrane) were able to integrate certain flux equations with the solution given as:

$$E_M = \frac{RT}{F} \ln \frac{(\mu_{Na^+})(C_{iNa^+}) + (\mu_{K^+})(C_{iK^+}) + (\mu_{Cl^-})(C_{oCl^-})}{(\mu_{Na^+})(C_{oNa^+}) + (\mu_{K^+})(C_{oK^+}) + (\mu_{Cl^-})(C_{iCl^-})} \quad (4-13)$$

where μ is the mobility of the cation and anion. For obtaining more meaningful results the mobility can be substituted by the permeability coefficient, P , which can be calculated from experimental data by the following:

$$P = \frac{f}{C_i - C_o} \quad (4-14)$$

C_i and C_o are the inside and outside concentrations in moles/liter, and f is the flux in moles/cm²/sec. If only the resting potential is considered, equation (4-13) may be simplified further by dropping out the Na^+ term since in this situation most cells have a permeability for Na^+ only about 1% that of K^+ . Furthermore, when the Na^+ concentration of the exoplasm is varied over a wide range, the resting potential of the cell is not affected, indicating that the Na^+ gradient across the membrane plays little if any role in the establishment and maintenance of the resting potential. In addition, it may be assumed that the resting potential of a nerve will reflect the ratio of the relative activities (or concentration in dilute solution) on the two sides of the membrane of the most permeable ion if this ion's permeability greatly exceeds all others. This would allow the use of the following modified Nerst equation:

$$E_M = \frac{RT}{nF} \ln \frac{C_o K^+}{C_i K^+} \quad (4-15)$$

If only action potentials are considered the final term need only show Na ion concentrations unless of course more accurate calculations are desired.

Chemical analysis of the major constituents in cyto- and exoplasm have provided numbers to be put in the various equations given above for comparing theoretical membrane potentials with experimentally determined ones (~90-100 mv). The best data obtained so far for resting potentials show differences between the two values, of ~30-40 mv, which is not surprising since the operation of the sodium ion pump (metabolic energy) is needed to actively extrude Na from the cytoplasm and maintain the difference in concentration of ions between the inside and outside of the cell. At the same time it should be emphasized that the polarized state is necessary for cell irritability, and therefore in order to maintain this state energy must be continuously expended. However, the fact that the correlations between observed and calculated membrane potentials have been even this close is an encouraging sign for the utilization of theoretical electrochemistry in attempting to analyze the significance of these potentials and, actually, is one step closer to understanding the nature of life itself.

4.4 Applications

In the preceding discussions an effort was made to relate various aspects of adsorption phenomena to biosensor operation. In considering the chemical senses of smell and taste it appears fairly certain that some form of selective adsorption takes place, which nearly simultaneously creates an electrical signal that triggers the nerve impulses to the brain. This entire mechanism is quite analogous to the recent development of gas and liquid chromatography which is widely used in analytical chemistry. In the case of gas chromatography, the most important link in the system is the detection of highly resolved, ultra low concentrations, of components as they evolve from a column containing a solid adsorbent, or a liquid adsorbent supported on an inert solid. When a liquid is employed, the difference in the partition coefficients for the gaseous components form the basis of the separation. When a solid adsorbent is used, it is the difference in adsorption coefficients which is important. Detectors which have been used in conjunction with chromatographic separations are found to include thermal conductivity, hydrogen flame, argon-beta rays, ionization gages, and radio frequency.

Another detector which has not been utilized to any extent up to the present time has a possible close relationship to the chemoreceptor mechanism. It utilizes the effect of the contact potential difference between two surfaces in electrical contact with each other. This contact difference of potential initially arises from a difference in the work functions of dissimilar surfaces, and in the laboratory is usually measured by the Kelvin-Zisman electrostatic technique of a variable or vibrating capacitor (Ref. 4-50). If a sufficient number of molecules, each of which is able to form some sort of a dipole with a surface atom, is adsorbed upon one of the electrodes, a change in electrical characteristics can be registered. This is due to a decrease or increase in the adsorbent work function, which may relate to the dipole moment, as follows:

$$\Delta \phi = \pm 4\pi \Delta (\sigma \theta \mu)$$

where

$\Delta \phi$ is the change in work function (ev.)

σ is the total number of sites per cm^2 which may be available to the adsorbate

θ is the fraction of σ which is available to the adsorbate, or the degree of monolayer coverage on the surface.

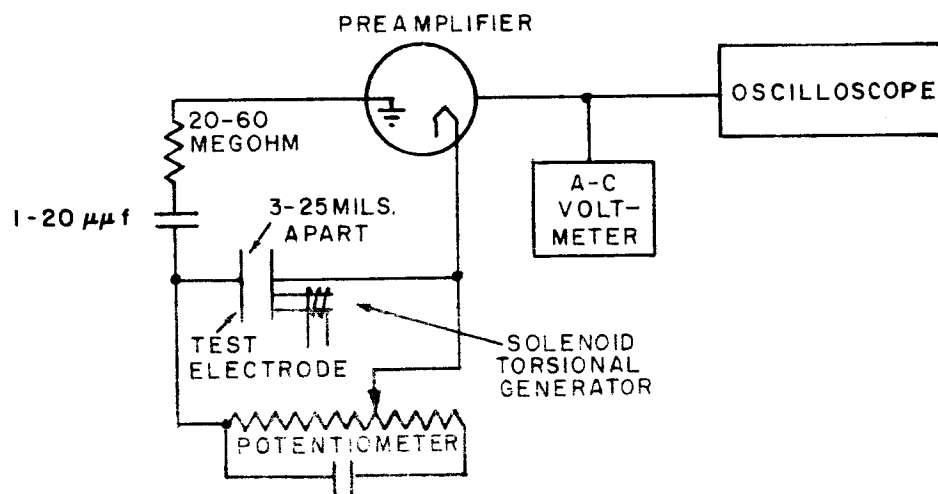
μ is the dipole moment per atom and is equal to the product of the electronic charge and the distance between the positive and negative layers.

At the present time, commercial chromatographic instruments are capable of detecting odors in concentrations of approximately one part per million, but research chromatographs appear to have been recently developed which may measure constituents present in ratios of the order of a billion to one. It is felt that some experimentation along the above lines might also offer a good possibility of developing a portable sensitive "black box" to simulate olfaction. The ultimate goal of this instrument would be to detect fantastically low concentrations ($< .01$ part per billion) of a specific odorous molecule in air. However, this latter goal will probably only be reached gradually (lower and lower concentrations) after some understanding is gained of the properties and peculiar adsorption characteristics of the specific molecules being investigated.

The initial experimental phase of this study would be to investigate the Kelvin-Zisman type measurement of the contact potential difference between two parallel plates separated by a very small distance. The set up simulates a condenser, the capacity of which varies periodically and sinusoidally when one electrode vibrates at a particular frequency and amplitude. The a-c component of the potential which is generated is fed to an amplifier and impressed on an oscilloscope. A potentiometer which is used to apply a dc potential between the electrodes is then adjusted; when the contact potential difference is compensated by the applied potential a variance in capacity between the test substrate and reference electrode produces no current flow and the sign wave deflection on the oscilloscope disappears. Apparently a number of different circuits can be used to obtain this null point. One such circuit which appears to be reasonable to utilize is shown on the simplified schematic, Figure 4-7.

FIG. 4-7

VIBRATING CONDENSER METHOD FOR MEASUREMENT
OF CONTACT POTENTIAL DIFFERENCE BETWEEN
TWO PARALLEL PLATES



Extreme precautions will of course have to be taken in the proper shielding of the circuit lest the noise level interfere with the desired results, since the ac voltages produced are only of the order of 0.1 to 1 microvolts.

The initial experiments can be carried out in a pure, dry atmosphere of N_2 , O_2 or a mixture of the two gases, wherein traces of various odorous organic molecules (substances with known vapor pressures are desirable to use as controls) can be introduced into the gas stream and identified by the change in contact potential which will occur due to either a physical or chemisorption process that will be different for each electrode. For the reference electrode (which will be vibrated by attaching it to a loud speaker driven by a variable oscillator) it would be desirable to use a material that would permit as small an amount of gaseous adsorption as possible on its surface. Clean polished steel, aluminum, gold or platinum which are coated with a thin film of teflon, calcium palmitate, stearic acid or some other substance may serve this purpose reasonably well.

There are various analogies between this "device" and the olfactory process in living organisms, since it is assumed that actual contact of odorous molecules with the sensory epithelium is necessary for olfaction to begin. If the initial response is caused by a chemical reaction at the gas-epithelium interface, a change in contact potential may be even greater than in the situation where only physical adsorption is occurring; also the mechanism of electron transfer, the nature of the bond formed, or the type of polarization involved is still not well understood when considering the creation of a dipole moment (or surface potential change) due strictly to physical adsorption. Various osmophoric groups in organic molecules such as nitro, methoxy, carbonyl, metcaptan, the number and proximity of the groups, introductions of double or triple bonds, length of side chain, "cis" or "trans" form of the molecule, open or closed rings, etc. can contribute substantially to the total dipole formation characteristics of various vapors. In any case, it must be emphasized again that we are probably dealing only with processes taking place in a partial monolayer (threshold requirements) which is not quite the same as if we were dealing with bulk chemical reactions. As previously indicated, it is quite likely that some sort of equilibrium condition in the adsorption process exists which depends upon parameters such as concentration, heat of adsorption, activation energies involved, if any, heat of desorption, and the height of the energy barrier which promotes or prevents diffusion onto the epithelium surface layers. A possible tie-in between the olfactory process

and the change in surface tension or pressure at the substrate surface has also been previously noted; this certainly offers another avenue for development. For example, a small shell trough filled with water can be set up on which a particular monolayer is spread. A bar or barrier laid across the trough behind the monolayer can be driven by a high-precision screw and made to push the leading edge of the film against a delicately suspended floating barrier which measures the pressure exerted by the film as it is compressed. Possibly, adsorption of as few as $10^5 - 10^7$ molecules per cm^2 may be sufficient to cause significant changes in the film-balance measurements.

In order to develop devices of maximum sensitivity and specificity it may be desirable to utilize the membrane sensor principle found in the olfactory epithelium.

In view of the olfactory epithelium being an aqueous mixture and colloidal solution of lipoids, proteins, and carotenoids (whether or not the actual cells are at the air interface) in addition to initial experimentation with various metallic, ionic, and organic solid surfaces, work must be done with gelatinous and liquid surfaces. Investigations should be made of partial monolayer adsorption, the possible interactions of osmophoric groups, and the ease of penetrability or dissolvment of various organic vapors at air-water, air-protein monolayer-water, air-carotenoid monolayer-water, etc. interfaces where this closer analogy to biological phenomena and the accompanying surface activity is present. One very important application of such devices would be the development of new methods and approaches for detecting the presence of living substances. For example, there is a strong likelihood that monolayers of various nutrients that can be ingested by living materials can be put down on a large number of substrates; with the deposition of a micro-organism on the test surface and the subsequent elimination, modification or decomposition of the nutrient monolayer, a very rapid change in contact potential or surface pressure will immediately occur, which should cease in less than one minute, when the metabolic reaction is completed. An application of this phenomenon can be easily seen in analogous experimentation on monitoring systems for germ warfare where as few as 1-2 pathogenic substances may be detected.

There are of course, numerous other potential applications for a specific, sensitive "molecule" detector which is being proposed in this re-

port, such as in intelligence or guerilla warfare operations where the possible identification and location of humans is desired, but may be very difficult to detect due to many factors, particularly distance and camouflage. Also, an individual may want to signal someone in order to relay vital information or signify that he is lost - the release of highly selective odorous molecules from a vial may be sufficient to do this, provided of course that the receiver has a detector which has the matching type of receptor substrate. The domain of food technology now stretches gas chromatography beyond its limits attempting to duplicate and identify various tastes and odors and is another area for investigation. It is not inconceivable that in the future various cancer detection probes can be developed from basic studies into the physical chemistry of membranes. In addition, the possibility exists of the utilization of biological membranes (or analogues) as possible device components. Air pollution monitoring of both open and closed atmospheres - ie, space cabin, etc. - could be immeasurably improved if more sensitive, specific and rapid methods for analysis became available.

The separation of numerous odorants into the seven primary odors (which have been noted earlier in this chapter), of ethereal, camphoraceous, musky, floral, pepperminty, pungent and putrid leads to an engineering concept of a smell warning device for spacecraft and airplanes whereby the pilot, through his olfactory organ, will perceive characteristic odors signifying the occurrence of a possible malfunction in an important operating unit of the system. This would be in conjunction with the standard warning system in current use. The extra few seconds of warning time gained could be quite critical.

The membrane sensor concept can be applied to develop transducers for a wide range of sensory modalities. The phenomenon of permeability changes in a living membrane frequently involving physiochemical reactions of varying complexity (analogous to olfaction, synaptic transmission, drug or poisonous action on nerves and muscle receptors, etc.) occur with most types of receptors -- mechanical (pacinian corpuscle, muscular contractions) thermal and photoreception, (rattlesnake pit, eye), olfaction, gustation, etc. Previously, the importance of monolayers in olfaction and nerve conduction was emphasized. This principle is easily expanded to include all types of receptors and a strong possibility exists that numerous types of monolayers,

and even complete artificial membranes can be put down on liquid (and gelatinous) surfaces to react to olfaction, light, radiation, thermal, mechano, etc. stimuli; thus developing new and better transducers based on principles of biodesign.

REFERENCES - SECTION 4

- 4-1 Moncrieff, R. W. J. *Physiol (London)* 130, 543, (1955)
- 4-2 Stuiver, M. "Biophysics of the Sense of Smell," PhD thesis, Groningen Univ., (1958)
- 4-3 Moncrieff, R. W., The Chemical Senses, Leonard Hill Ltd. London (1951) p. 34.
- 4-4 Briggs, M.H. and Duncan, R.B., *Nature*, 191, 1310, (1961)
- 4-5 Zechmeister, L., *Chem. Revs.*, 34, 267, (1944)
- 4-6 Moulton, D. L., *Nature*, 195, 1312, (1962)
- 4-7 Muller, A., *Perfum. Essent. Oil Rec.*, 27, 202, (1936)
- 4-8 Timmermans, J., *Nature*, 174, 235, (1954)
- 4-9 Amooore, J.E., *Perfum. Essent. Oil Rec.*, 43, 321 (1952)
- 4-10 Moncrieff, R.W., *J. Physiol (London)* 130, 543, (1955)
- 4-11 Davies, J.T. and Taylor, F.H., *Nature*, 174, 693 (1954)
- 4-12 Davies, J.T. and Taylor, F.H., "Molecular Shape, Size and Adsorption in Olfaction," Proc. 2nd Annual Congress of Surface Activity, Vol. 4, Academic Press, N.Y. (1957)
- 4-13 Amooore, J.E., *Proc. Sci. Sec. Toilet Goods Assoc. No. 37*, Suppl. I, 13, (1962)
- 4-14 Amooore, J.E., *Nature*, 198, 271, (1963)
- 4-15 Gesteland, R.C., Lettvin, J.Y., Pitts, W.H. and Rojas, A. "Odor Specificities of the Frog's Olfactory Receptors," in Olfaction and Taste. Zotterman, Y., ed. by MacMillan Co. N.Y. (1963)
- 4-16 Pauling, L., *Chem. Eng. News*, 24, 1064, (1946)
- 4-17 Beets, M.L.J., *Naturwiss.*, 47, (19) 56, (1960)
- 4-18 Witheridge, W.N. and McCord, C.P., Odors, Physiology and Control. New York (1949)
- 4-19 Mateson, J., *Ann. N.Y. Acad. Sci.* 58, 83 (1954)
- 4-20 Adrian, E., *Acta Physiol. Scand.* 29, 5, (1953)
- 4-21 Adrian, E., *J. Physiol (London)* 100, 459, (1942)
- 4-22 Adrian, E., *J. Physiol (London)* 112, 38, (1950)
- 4-23 Wright, R.H., *J. Appl. Chem.* 4, 611, (1954)
- 4-24 Wright, R.H., Reid, C. and Evans, H. L. V., *Chem. and Indust.* 37, 973, (1956)

- 4-25 Wright, R.H., *Nature*, 190, 1101, (1961)
- 4-26 Wright, R.H., *Nature*, 198, 455, (1963)
- 4-27 Briggs, M.H. and Duncan, R.B., *Nature*, 191, 1310 (1961)
- 4-28 Dyson, L.M., *Nature*, 173, 831 (1954)
- 4-29 Moelwyn-Hughes, E.A., *Physical Chemistry*. Pergamon Press, London, (1957) p. 183.
- 4-30 Beck, L.H. and Miles, W.R., *Science*, 106, 511 (1947)
- 4-31 Ottoson, D., *Acta Physiol. Scand.* 35, 122, (1956)
- 4-32 Dravnieks, A., *Nature*, 194, 245, (1962)
- 4-33 Ryuji, A., *J. Phys. Soc., Japan*, 11, 104, (1956)
- 4-34 Davies, J.T. and Taylor, F.H., *Nature*, 174, 693, (1954)
- 4-35 Davies, J.T., Biological Receptor Mechanisms, A Symposium. Academic Press, N.Y. (1962) p. 170.
- 4-36 Davson, H. and Danielli, J.F., *J. Biochem.* 32, 991, (1938)
- 4-37 Skramlik, J., *Pflug Arch. Ges. Physiol.* 249, 702, (1948)
- 4-38 Morimura, S., *Tohoku J. Exp. Med.* 22, 417, (1934)
- 4-39 Backman, E.I., *J. Phys Path. Gen.*, 17, 1, (1917)
- 4-40 Dikstein, S., Biophysics of Physiological and Pharmacological Actions. Amer. Assoc. for the Adv. of Science, Wash. D.C., (1961) p. 49.
- 4-41 Shanes, A.M., *Pharmacol. Revs.*, 10, 59, (1958)
- 4-42 Shanes, A.M., *Proc. of the Int. Union of Physiol. Sciences*, Vol. I, No. XXII, 586, (1962)
- 4-43 Mullins, L.J., *Chem. Revs.* 54, 289, (1954)
- 4-44 Swift, M.R., London, H.P. and Vander Kloot, W.L., *Proc. Natl. Acad. Sci. U.S.* 46, 1415 (1960)
- 4-45 Meyer, K.H., *Helv. Chim. Acta.* 20, 634, (1937)
- 4-46 Archer, R.J. and LaMer, V.K., *J. Phys Chem.* 59, 200, (1955)
- 4-47 Blank, M., *J. Phys. Chem.*, 65, 1698 (1961)
- 4-48 Donnan, A., *Electrochem.*, 17, 572, (1911)
- 4-49 Hodgkin, A.L. and Katz, B., *J. Physiol (London)* 108, 37, (1949)
- 4-50 Zisman, W.A., *Rev. Sci. Inst.* 3, 369, (1932)
- 4-51 Healer, J., Arms, A., Merchant, J. and Britt, E., A Summary Report on a Review of Biological Mechanisms for Application to Instrument Design. ARA 1025, Allied Research Associates, Inc. January, (1962)

5. Thermoreception

As a part of the complex of the sense of touch, the thermal sense has been studied for centuries by means of the introspective method, in which the attempt is made to subjectively analyze sensations into their presumed elementary components (Ref. 5-1). Although some interesting insights into the relationships between the physical stimulus and sensation were derived from such studies, quantitative information on the detailed aspects of the relationships became available only within the past 40 years, mainly as a result of the development of electrophysiological techniques, particularly the cathode ray oscilloscope. Because of the complexity of biological materials, which pose many difficult technical problems of experimentation and analysis, the data on thermoreceptors is still far from complete. However, as the following review of the relevant aspects of the available data will show, the ingenious methods by which animals detect temperature effects, and the way this information is processed, indicate a realm of interaction between the technology of nature and its physical counterpart. This convergence and mutual reinforcement at both the technical and conceptual levels between the biological and the physical approaches to essentially the same problems can come about only through a broad but closely integrated interdisciplinary program such as that under which the present comprehensive study has been organized.

5.1 Warm and Cold Spots

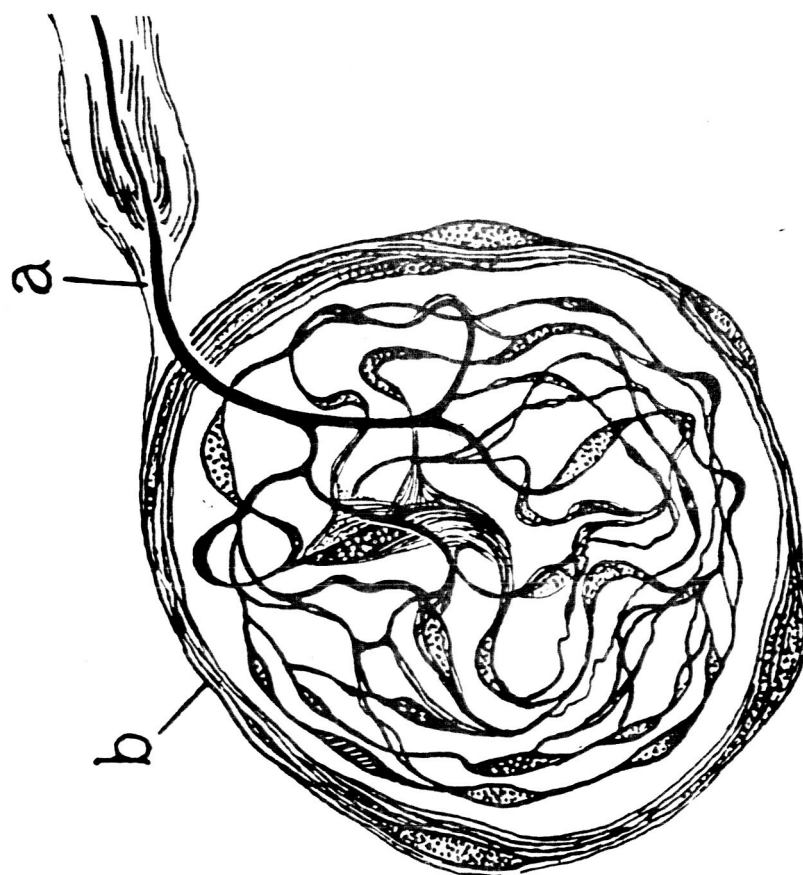
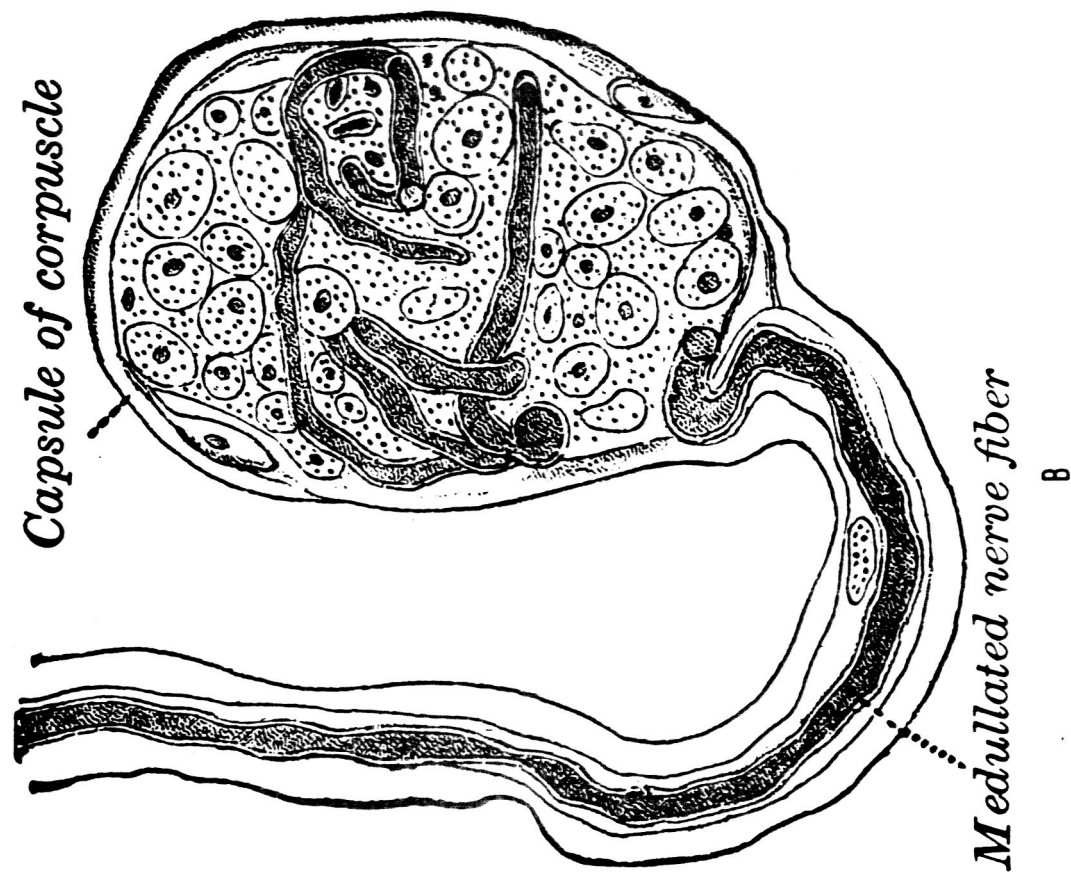
The scientific period in the study of thermal biotransducers may be considered to begin with the significant discovery by Blix (Ref. 5-2) of the cutaneous "warm" and "cold" spots in man. Using mild electric shock delivered by a small electrode, in order to stimulate with a non-specific, point source of energy, he found separate spots that, when stimulated, produced sensations of warmth and cold, as well as of pressure and pain. He later confirmed the independence of warm and cold spots by using as his stimulator a hollow conical metal point through which warm or cold water could be passed. Warm and cold spots were reported independently about the same time by Goldscheider (1884), and by Donaldson (1885).

A large amount of work followed this basic discovery. An obvious follow-on is to determine the distribution of warm and cold spots, and the

whole human body has been explored millimeter by millimeter (Ref. 5-3 through 5-11, also see Hensel Ref. 5-12 for review). Cold spots, which are most numerous in the face (16-19 per cm²), outnumber the warm by 90:1. Another obvious thing to do is to look for the neural receptors which mediate the warm and cold spot sensations. The technique is to first find a spot, then examine it microscopically. Unfortunately, the results are not as clear as could be desired. Donaldson (1885) and Goldscheider (1884) excised the areas under cold and warm spots but could find only free nerve endings. Negative results were also reported by Haggqvist (1913), Dallenbach (1927), and Pendleton (1928), (see also Ref. 5-13 and 5-14). Why conventional histological methods fail to show any such relationships is not clear. On the other hand, Strughold and Karbe (Ref. 5-10), found a good topographical correlation between the cold spots and a type of nerve ending called the "Krause end bulb" (described below). They first mapped the cold spots in the periphery of the cornea of the eye, then dropped methylene blue tissue stain into the eye to make nerve endings visible under the corneal microscope. Similarly, Bazett et al (Ref. 5-15), studying the foreskin of the penis, injected methylene blue intra arterially, and although they found seven different types of end organs, the correlations between the distribution of cold spots and Krause end bulbs, and warm spots and Ruffini end organs, was considered to be good. Unfortunately, it does not appear that these workers performed the obvious test of first locating an end bulb, and stimulating it under microscopic control. If a cold or warm sensation invariably followed, they would have a strong case.

5.2 Warm and Cold Receptors

The idea that Krause end organs (Fig. 5-1) are the receptors for cold, and Ruffini end organs (Fig. 5-2) mediate warmth, originated with von Frey (1895), an eminent turn of the century physiologist (Ref. 5-16). He assigned the Krause bulbs to cold because they occur in the conjunctiva of the eye (at the edge of the cornea) and the glans penis, which are insensitive to cutaneous pressure. The conjunctiva is also insensitive to warmth. He assigned the Ruffini endings, which are plentiful in the finger pulp and less so in the skin, to warmth. Because of von Frey's prestige, this view is repeated uncritically even today. However, the weight of evidence does not support it. von Frey was, as Boring says, "doing little more here than delegating the last unassigned organ to the last available sense" (Ref. 5-1). Gilmer in 1942, cites evidence



a, Sensory nerve fiber; b, capsule. From Ranson S. W., and Clark, S. L.: *Anatomy of the Nervous System*, W.B. Saunders Co., 1959.

FIG. 5-1
Examples of Krause's End Bulb

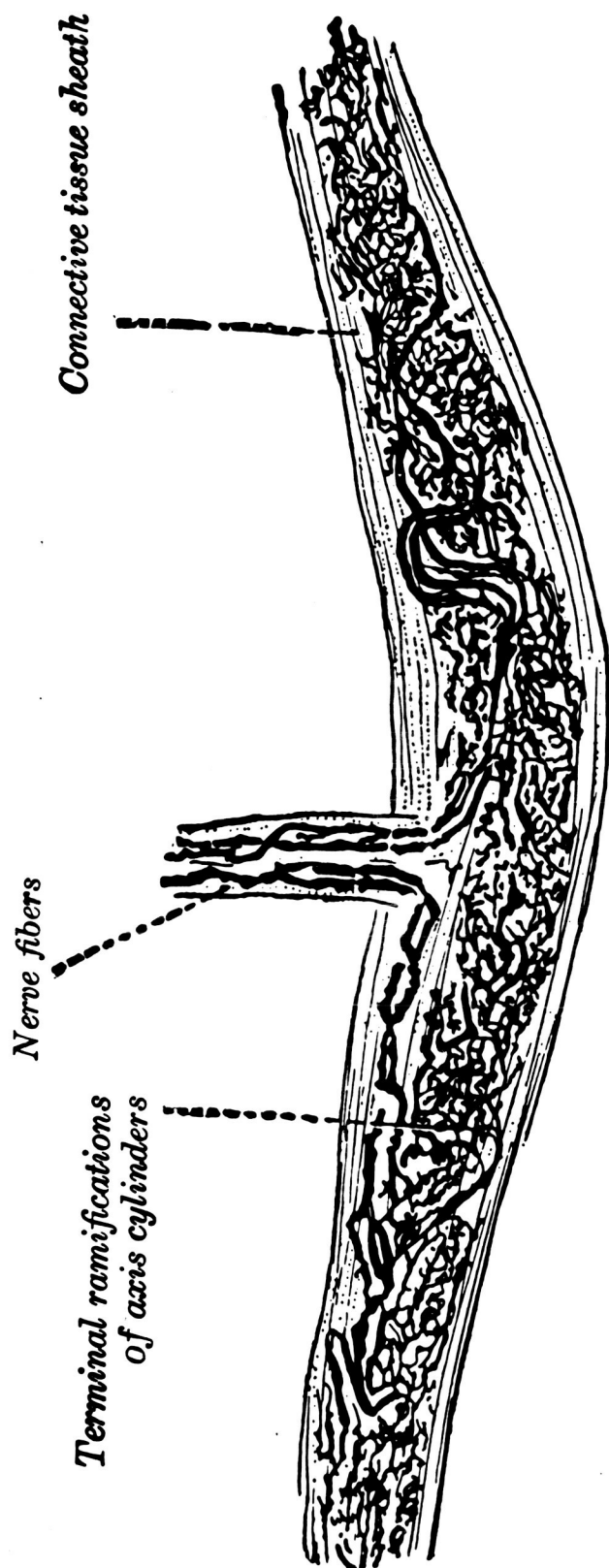


FIG. 5-2
Nerve ending of Ruffini. (After A. Ruffini.)

that seems to implicate the so called "glomus bodies", nerve endings reminiscent of simple Krause organs, in temperature reception. However, even less is known about the structure and distribution of these bodies.

However, certain other evidence seems to implicate the Krause and Ruffini end organs. The reaction time for warmth is consistently longer than that for cold, suggesting that the warm receptors should be located deeper in the skin than the cold receptors (Ref. 5-7, 5-16 through 5-20). More direct evidence is provided by Bazett et al (Ref. 5-15), who stretched the foreskin of the penis out to form a flat sheet so that the sensitive spots in one layer of skin could be stimulated from either side of the double fold. The rate of transmission of the temperature wave through the fold of skin was determined by thermoelectric recording. The value obtained, about 1 mm/sec., has recently been confirmed by Hensel and Zotterman (Ref. 5-21). By relating this transmission time to the reaction time of the subject, the average depth of cold receptors was calculated to be 0.17 mm, which is in good agreement with the average depth of Krause end organs. Warm receptors are calculated to be about 0.3 mm deep, a figure in the range where Ruffini endings are numerous. Against such circumstantial evidence is the fact that these end organs are not found in many areas of the skin where cold and warm spots are found.

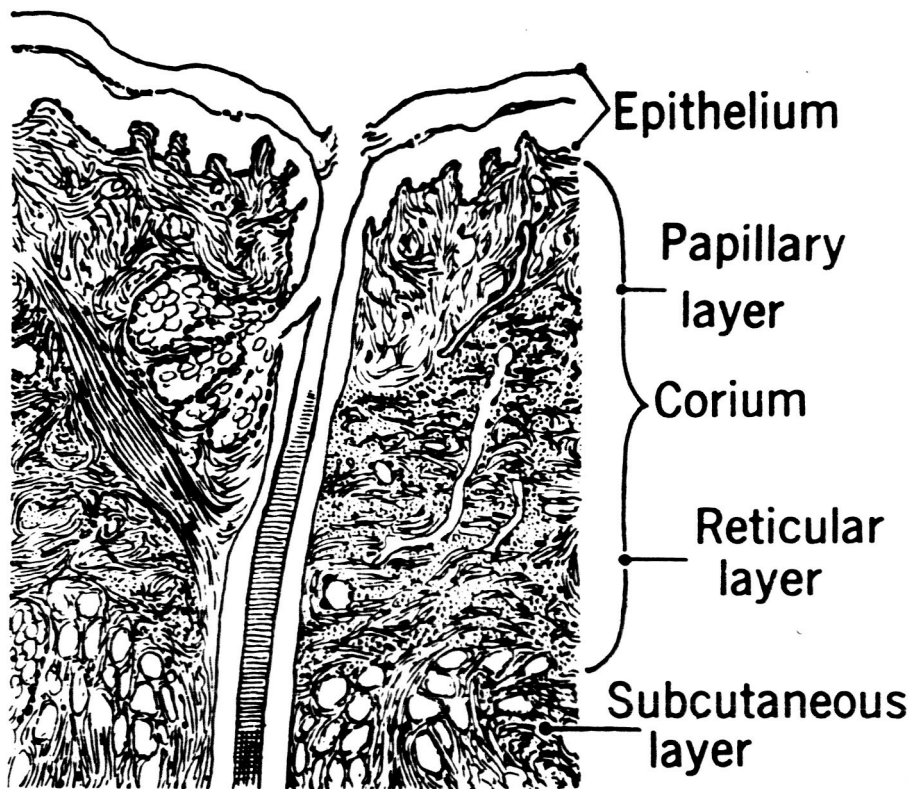
One should not get the impression that these end organs are of a single uniform type. The Krause end bulbs (Fig. 5-1) are cylindrical or oval capsules formed by the expansion of the connective tissue sheath of a medullated nerve fiber, and has a semifluid core in which the axis-cylinder of the nerve terminates either in a bulbous extremity or in a coiled up plexiform mass. The simplest and smallest are found in the conjunctiva of the eye. In man they are spheroidal in shape, but in most other animals they are cylindrical. The largest forms of these end-bulbs are plentiful in the glans penis and clitoris. These "genital corpuscles" are constricted by connective-tissue septa into from two to six knob-like masses. End bulbs of various forms have a wide distribution, being found in the epiglottis, nasal cavity, peritoneum, the mucous membrane of the lips and tongue, the lower end of the rectum, and in the epineurium of nerve trunks, that is, the connective-tissue surrounding the nerve bundle. In the tendons, ligaments and synovial membrane of certain joints, such as those of the fingers, the end bulbs are rounded

or oval. Krause end bulbs are thus found in some unlikely places for a cold receptor. In view of the diversity in shape, size and locations, the term should be understood to refer not to a single type of organ with a single or special function, but to a general class of organ configuration. The case for Ruffini corpuscles (Fig.5-2) being the warm receptors is even less secure. They are elongated bodies of considerable size whose capsule encloses several bundles of connective-tissue fibers. Several nerve fibers enter the corpuscle and ramify extensively, the branches lying between and partly encircling the small connective tissue bundles. The nerve branches show varicosities and end in small free knobs. They are found in the human finger, situated principally at the junction of the corium with the subcutaneous tissue (Fig.5-3).

Cold and warm spots can be stimulated by mechanical (Jones, 1940) and even electrical stimulation, as Blix (1882) did in the original experiments. The reader can find cold spots on the back of his own hand by exploring with the tip of a sharp pencil. Chemicals such as menthol can also produce sensations of cold. These are taken as evidence for the specificity of the receptors.

The uncertain identification of specific temperature sensors suggests that what is specific is not the transducer, but the part of the brain to which it is connected. It is known that C-fibers (small, unmyelinated nerve fibers) fire to warming, and A-fibers (large diameter, myelinated nerve fibers) to cooling. Cutaneous nerve fibers with endings of various types, free and encapsulated, may fire to temperature changes, but only those connected to the temperature area of the cortex are thought to produce the sensations of cold and warmth. This view is based on the more general idea that the nervous system is mode specific, the sensory modes being sorted out topographically at the brain. The specific sensory endings, such as the retina or a Pacinian corpuscle, then serve only to efficiently couple a particular type of physical energy system to the nervous system. The neural transmission system appears to be sorted out qualitatively, so that the neural pulse code does not contain information of this type, but of the pattern of the intensity of a band of energy impinging on the sensor. In the case of temperature, the nerve fibers are probably sorted according to the area of the body innervated, the pulse code of a particular fiber conveying information on the temperature of a particular area.

FIG. 5-3
Corium



Section through skin to show structures in the corium (or derma)

There is some electrophysiological evidence in cats that certain cells in the somesthetic cortical area respond specifically to cooling or warming of the tongue (Ref. 5-22 and 5-23). Some cells respond only to thermal stimulation, while others were activated by mechanical or gustatory stimuli as well. There is also evidence in the cat demonstrating specific sensory endings which fire only to cooling or warming, and not to mechanical or other stimulation (Ref. 5-12 and 5-24 through 5-27). In these experiments the discharges to cooling of the tongue were recorded from the nerve leading from the tongue. However, the sensory receptor does not appear to have been identified histologically.

One could expect to find, due to normal variations, even in a group of fibers connected to the temperature area of the brain, that some endings are more efficient than others, and even to see some marked morphological differences. Moreover, thermoreception is not a single unitary function, but consists of at least two systems, each occurring at different neural levels. Temperature sensation is one function, and the activation of central temperature regulating mechanisms is another. Each is served by its own sensor transmission lines. This further complicates a very complex subject.

There are some recent data on the cat's tongue relating temperature changes to neural discharge rate, recording from the cold fibers of the lingual nerve and the warm fibers of the chorda tympani (Ref. 5-27 through 5-29). Both cold and warm fibers fire at a high frequency at a temperature change (see Fig.5-4), then settle to a slower firing rate that is related to the temperature. Both types of fibers also fire at the highest rate under conditions of constant temperature within certain preferred temperature ranges (see Fig.5-5). These are the characteristics of 'tonic' neurons, that fire at a level somewhat proportional to some function of the intensity of the stimulus. ('Phasic' neurons fire at a high frequency at the moment the stimulus is applied, then stop firing or revert to the pre-stimulation level even though the stimulus continues.)

5.3 The Thermal Stimulus

A question which appears to be in the process of being settled is that of the adequate stimulus for the cutaneous temperature sensors. Normally, there is a temperature gradient between the circulating blood (37.5 degrees C) and the skin surface (32-33 degrees C on the forearm). The adequate stimulus

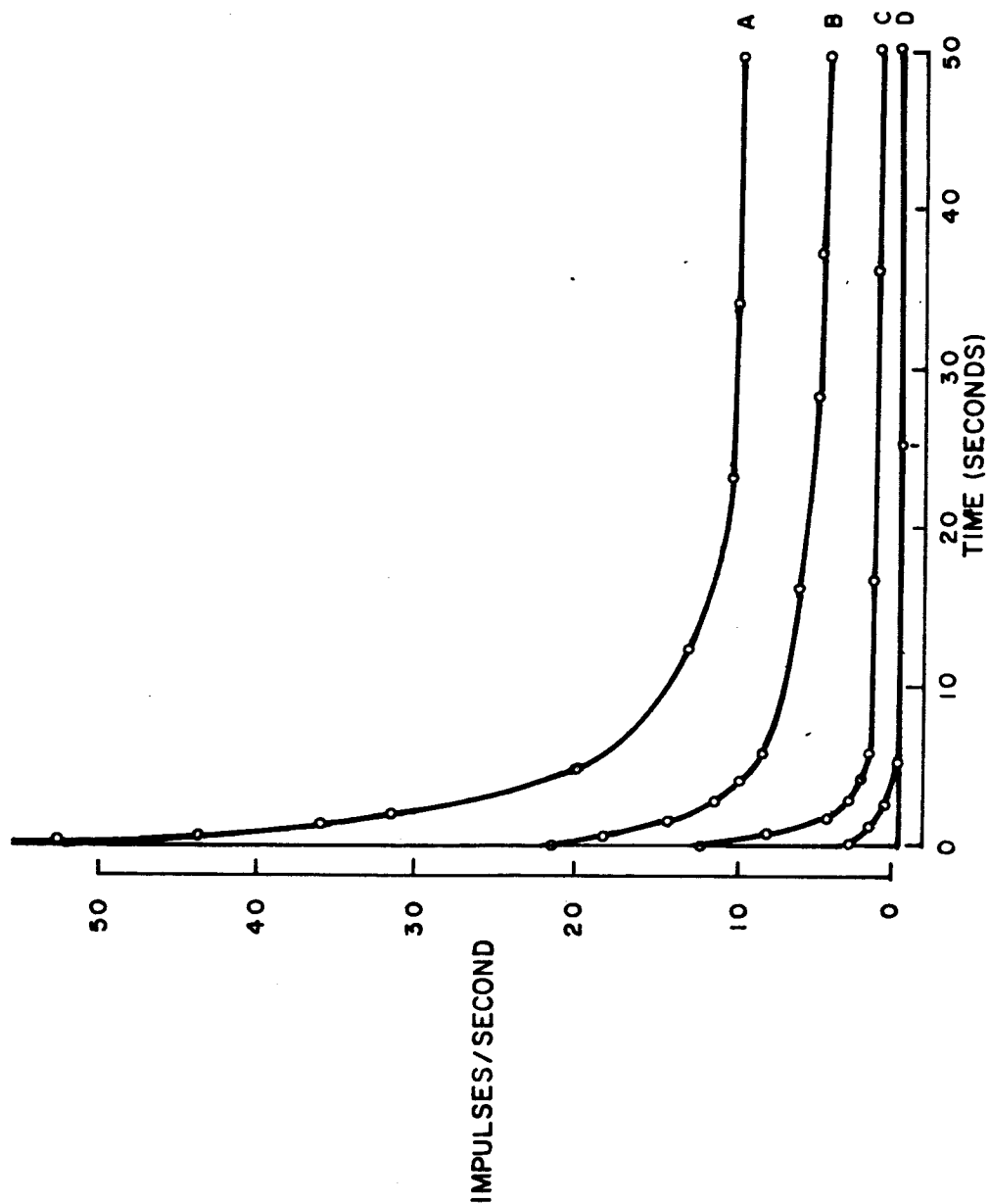


FIG. 5-4

IMPULSE FREQUENCY OF A SINGLE COLD FIBER IN THE CAT LINGUAL NERVE
DURING INDUCTION OF RAPID TEMPERATURES IN THE TONGUE

A, FROM 32° TO 30° C; B, FROM 40° TO 38° C;

C, FROM 42° TO 40° C; D, FROM 44° TO 42° C;

COOLING STARTS AT TIME ZERO. [FROM HENSEL & ZOTTERMAN (5-25)]

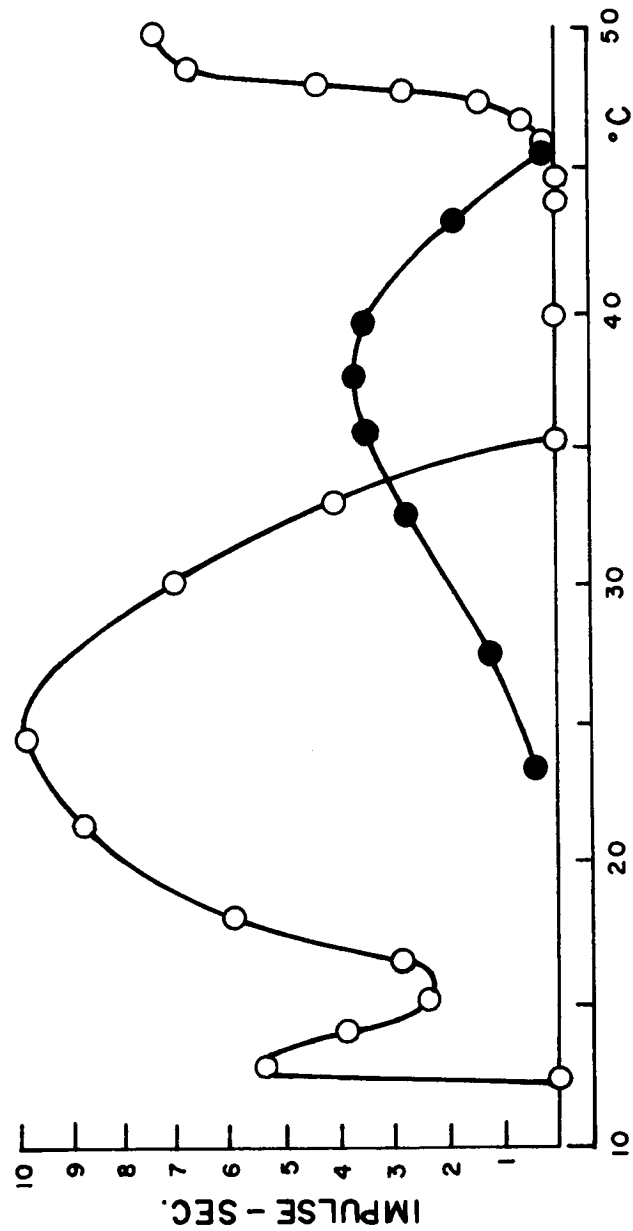


FIG. 5-5

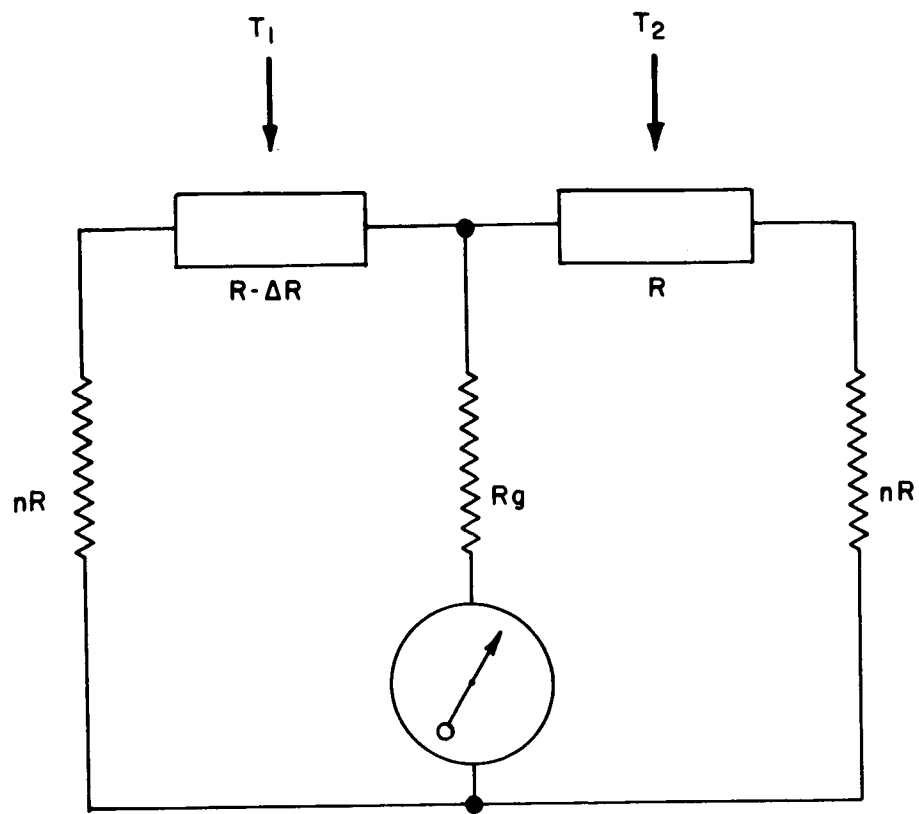
GRAPHS SHOWING IMPULSE FREQUENCY OF THE STEADY DIS-
CHARGE OF A SINGLE COLD FIBER (OPEN CIRCLES) AND OF A SINGLE
WARM FIBER (FILLED CIRCLES) AS A FUNCTION OF THE TEMPERATURE
OF THE RECEPTORS WITHIN THE RANGE OF 10 TO 50° C. [FROM ZOT-
TERMAN]

might be a disturbance of the gradient as such, or a change in temperature at the receptors, with the gradient shift purely incidental. A recent restatement of the first view was proposed by Lele et al. (Ref. 5-30), in which non-specific sensors in different strata of the skin operate as a thermopile-type of bolometer (Fig. 5-6). The major difficulty with this view is that it assumes that qualitative information is coded by nerve impulses. As was observed before, this is not likely since it appears that sensory fibers are sorted qualitatively, and transmit only power spectrum information to the specific receptive field in the brain. There are several lines of evidence showing that gradients are less important than temperature per se. At the behavioral level, Bazett and McGlone (Ref. 5-31) observed that cooling from the lower surface of the double fold of the skin of the prepuce of the penis led to a sensation of cold in the upper surface. From experiments in which the temperature change and the change in the spatial gradient ran in opposite directions, Hensel (Ref. 5-32) concluded that the simple warming or cooling of the receptors, rather than the intracutaneous temperature gradient or its temporal change, was the effective stimulus. Hensel and Zotterman (Ref. 5-21) recorded the action potentials from the cold fibers of the lingual nerve of the cat when cold stimuli were applied to the tongue to produce negative or positive intracutaneous temperature gradients. The nerve preparations chosen were those containing cold fibers supplying only the upper surface of the tip of the tongue. Cooling the upper surface resulted in the firing of the cold fibers; warming had the opposite effect. On cooling the lower side no impulses appeared at first, but within a few seconds, when the cold had penetrated the tongue and reached the upper surface, discharges occurred with increasing frequency. In other experiments the receptor layer was cooled by injecting cold solutions into the lingual artery, with similar results. Such experiments indicate that, at least for cold receptors in the cat's tongue, the direction or slope of the intracutaneous temperature gradient is not the decisive stimulus, but the temperature itself is.

5.4 Central and Other Influences on Sensor Performance

Another difficulty in a precise engineering study of biological transducers is that biological systems are dynamic and highly selective with regard to specific inputs. A stimulus is not injected into a passive system, as is essentially the case in physical systems. In a biological organism, the stimulus

FIG. 5-6
BOLOMETER



is immediately modified by the on-going activities of the nervous system (C. F. Brazier, et al, 1961). It is generally assumed that the way to reduce on-going activity, which in effect constitutes noise, is to anesthetize the organism. While it is true that the background neural activity becomes greatly reduced, such a preparation cannot be responding to and processing stimuli as the normal "awake" organism does. Moreover, all organisms are controlled by motivational factors, as well as by the external stimulus conditions. The effect that a given stimulus produces in an organism is not solely a function of its physical parameters, but to an important degree, on how the animal feels or what it is doing at the time. For example, the threshold to cutaneous stimulation is markedly affected by how hungry the animal is, as shown by Cheng (Ref.5-33), who used cutaneous stimulation as the conditional stimulus for food in dogs. The stimulation threshold was markedly lower, the longer the animal had been deprived of food (up to a point). Electrophysiologically, discharges can be picked up from receptors even under constant ambient conditions. These are presumably injected into the system by central thermoregulatory activity (Ref.5-34 and 5-35). These show that the sensitivity and selectivity of the sensors are modified by centrifugal impulses from the brain itself (cf. Livingston 1959, for review). In an anesthetized animal, motivational factors are probably minimal, or at least are an unknown quantity; sensory thresholds and other characteristics as determined in such preparations, may, therefore, have no relationship or relevance to what happens in the normal animal. Finally, it may appear that the determination of thresholds by electrophysiological means, using the frequency of firing as the criterion, permits an objective determination of thresholds and other characteristics, but such a criterion is arbitrary, since the relationship to what the animal senses has not yet been established. These considerations indicate some of the complex interactions that occur even in a seemingly simple function like thermodetection. The fact that organisms function effectively with these complexities is indicative of data processing capabilities astonishing in their sophistication.

5.5 Radiant vs. Contact Stimulation

There are two puzzling differences between the effects of radiant and contact stimulation in man. First, warmth cannot be sensed at all when small areas are stimulated with radiant energy. In areas of less than about 700 mm² only pain is experienced when the radiation becomes strong enough to elicit any sensation at all (Ref. 5-36). In contrast, warm spots may be stimulated with a contact area of 1 mm² or even less. Secondly, positive summation of the simultaneous radiant stimulation of two separate areas is quite marked. This means that the radiant energy per unit area required to produce a threshold sensation is much less when two separate areas are stimulated simultaneously than when either is stimulated alone. That central (brain) factors are involved is indicated by the fact that summation occurs when the backs of the two hands are stimulated; however, there is no summation when a hand and the forehead are stimulated together. In contact stimulation, the spatial summation of effects is at best slight, and in some cases even negative and when two adjacent areas give equal intensity scores when stimulated, stimulating them simultaneously results in only a slightly higher score. When two areas have markedly different scores, stimulating them simultaneously gives a lower score than the maximum singly, and only slightly higher than the average of the two. This phenomenon recalls that of lateral inhibition of response in the Limulus eye, when two adjacent receptors are stimulated simultaneously (Ref. 5-37). It would be significant if a similar mechanism is operative in the cutaneous temperature sensor system. A program to investigate this and related questions at the behavioral and electrophysiological levels would yield important information on the nature of the transmission system of the cutaneous thermal receptors.

The preceding results have raised the possibility that there may be two different thermal receptor systems, or else radiant and contact stimulation act differently on the same receptor systems. The latter idea seems difficult to accept. Since the receptors lie below the upper layers of the skin, the temperature at the receptors would seem to be independent of the manner in which it is produced. On the other hand, as was mentioned earlier, it seems possible for a variety of nerve endings, located in various strata of the skin, to transmit thermal information.

This review has indicated that the information currently available on the structure and properties of cutaneous thermal receptors is, by engineering requirements, at a fairly rudimentary level. It has also pointed out some of the problems that stand in the way of a proper engineering study; these stem basically from the extreme complexity of the biological system. In such a circumstance, it is reasonable to study as simple a case as is compatible with the engineering aims, and this has been done to a certain degree, in the infrared sensor of the snake, which will be described below. But first it is necessary to describe what is known about the internal thermal control system, which, in homothermic organisms, normally maintains the body temperature within a few degrees of the basal level, despite extremely variable ambient conditions.

5.6 Thermoregulation in Homothermic Animals

Bodily thermoregulation has been studied for over 70 years, and many reviews are available (Ref. 5-38 through 5-50). The basic facts seem to be that as yet unidentified thermoreceptive structures located in the highly-vascularized supraoptic and preoptic areas of the hypothalamus (Ref. 5-51) continuously monitor the temperature of the blood, and send appropriate signals to effector systems which actually warm or cool the blood (Ref. 5-52 through 5-58). The effector systems and reactions available for this purpose include pulmonary ventilation, cutaneous blood flow, sweating, salivation, piloerection, skeletal muscular activity such as shivering, movement of inter- and intra- cellular water, and use of the limbs and body positions. As would be expected, the central thermoregulatory mechanism is basically a closed-loop servo system (see Figure 5-7). The remarkable stability of the temperature regulating system appears at present to be an extremely complex problem for analysis.

It is not yet known whether the hypothalamic thermodetectors are sensitive both to warming and cooling. Since the discharge of single nerve units of the thermodetectors has not yet been recorded, it is not known whether the temperature rise that activates the heat-loss mechanisms produces an increase or a decrease, or both, in firing frequency. It will be remembered that a cutaneous thermoreceptor unit fires maximally at an individually characteristic temperature (cf. Figure 5-5). This temperature varies considerably

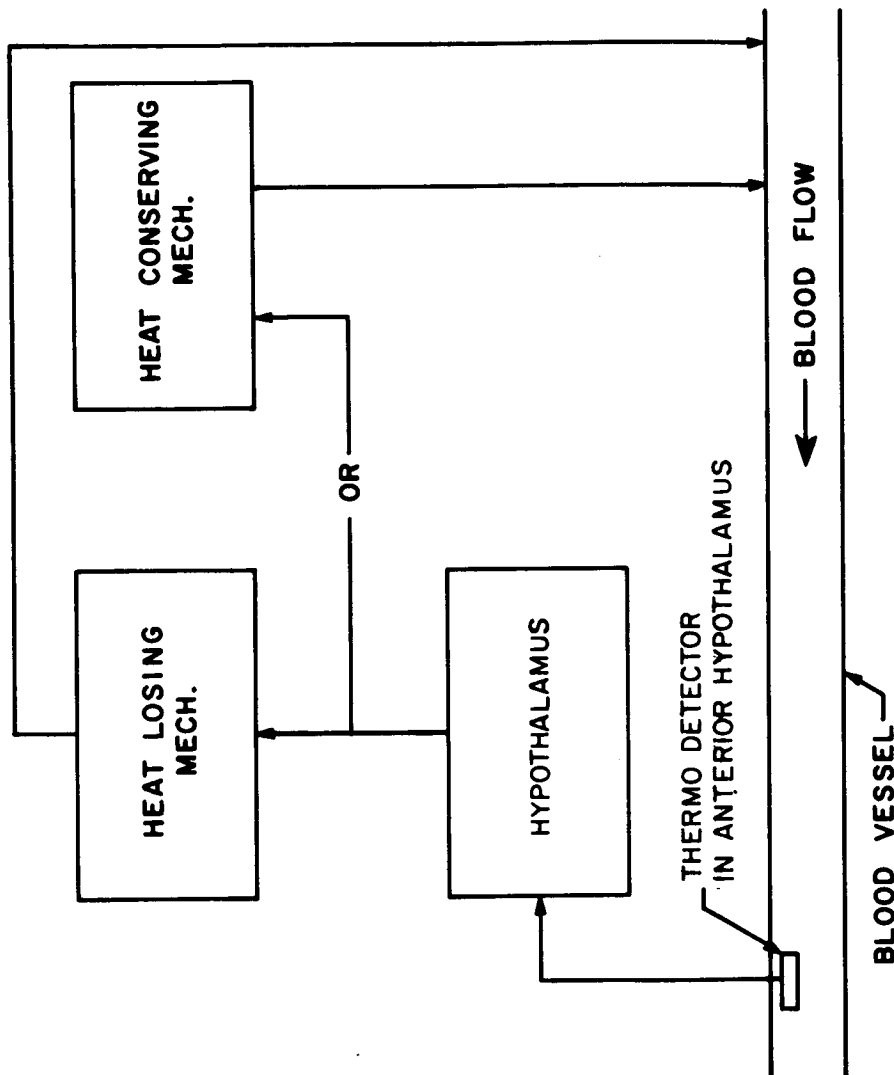


FIG. 5-7
 HYPOTHETICAL ANATOMICAL CONNECTIONS
 OF THE CENTRAL THERMOREGULATORY SYSTEM
 (A CLOSED LOOP SERVO SYSTEM)

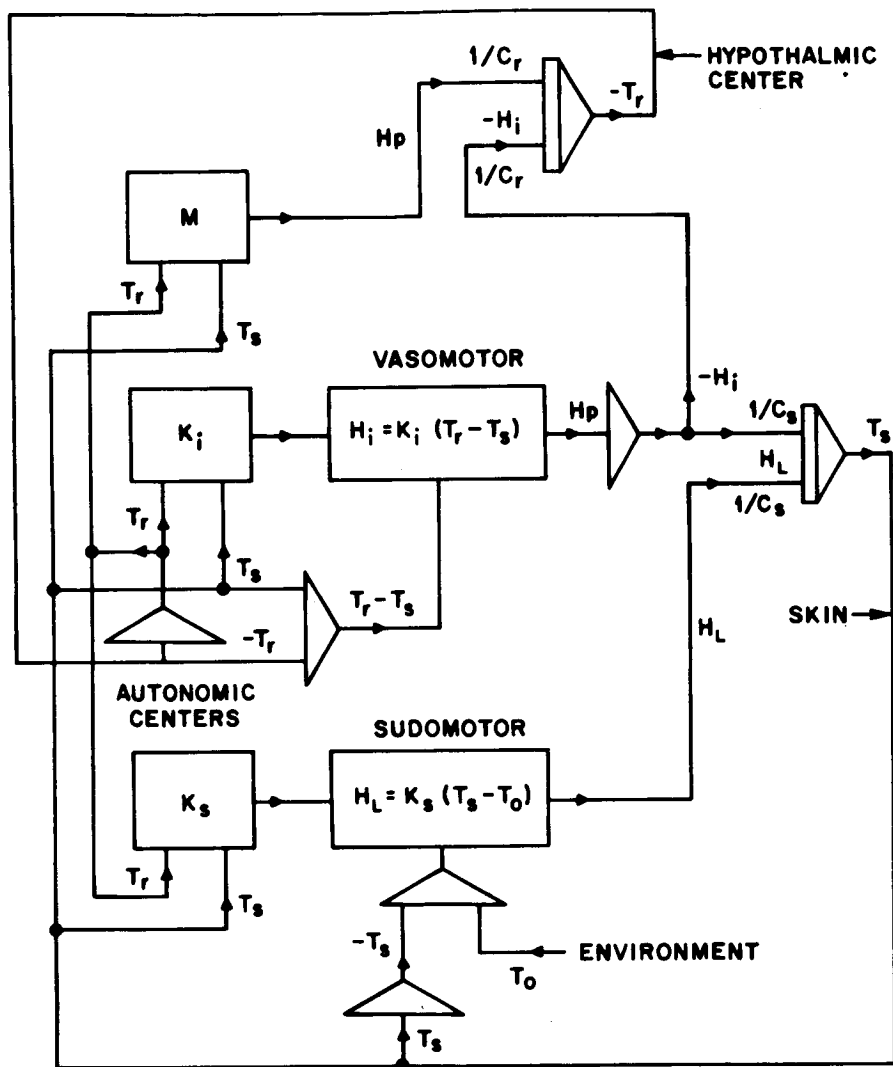
within the receptor population. A temperature change therefore produces deactivation of some receptors and activation of others. Those which fire above the normal range of body temperature may be classified as warm receptors, and those below as cold (note, in contrast, that stimulation of cutaneous warm or cold spots produces sensations of warmth or cold). The only well-known property of the hypothalamic thermodetectors is that the range of temperatures within which the effector systems show prominent reactions is relatively narrow, extending from about 1 degree C below to a few degrees above the normal brain temperature (Ref. 5-57 and 5-59). Dynamic range is secondary to the requirements for fine resolution. A curious feature is that at brain temperatures above 41 or 42 degrees C, a reversal of effector reactions can occur, i. e., the feedback switches from negative to positive. It may be speculated that this may be a built-in destruct mechanism. The only electrical activity recorded from the anterior hypothalamus in response to a change in the temperature of the blood (warming) has been a slow potential change of 0.5 to 1.0 millivolt per 0.1°C (Ref. 5-60 and 5-61). A similar local potential field develops in the medulla oblongata in the brain stem when the arterial pCO_2 is changed. (Ref. 5-62.)

The stability of the central thermoregulatory system in the face of the variations in body temperature that can occur not only between widely separated superficial areas, but also between surface and deep regions, has been noted before. The understanding of the way regulation is achieved without undue oscillations or other perturbations in complex biological organisms may provide systems engineers with new insight. On the other hand, many techniques by which perturbations can be controlled are already known to engineers, at least in linear systems. The thermoregulatory system operates non-linearly, except possibly for ranges of small changes. The problems of theoretical analyses can therefore become extremely complex. However, a systems study of central biological thermoregulation may well be worthwhile.

A computer model of the thermoregulatory system, based on assumptions of linearity, appears in Figure 5-8. This model also assumes that the cutaneous thermoreceptor system is an integral, anatomical part of the central thermoregulatory system. However, recent studies Benzinger (Ref. 5-63) on human subjects indicate that this need not be the case. Instead, Benzinger suggests that there are two independent thermoregulatory systems, one peripheral and

FIG. 5-8

A COMPUTER MODEL OF THE THERMOREGULATORY SYSTEM. (FROM REF. 5-70)



M = HEAT GENERATOR (BATTERY)

H_p = HEAT PRODUCTION

T_r = RECTAL TEMPERATURE

T_s = SKIN TEMPERATURE

T₀ = AIR TEMPERATURE

H_i = HEAT TRANSFERRED FROM CORE TO SURFACE

H_L = HEAT LOSS FROM SURFACE OF SKIN

K_i = FUNCTION OF VASOMOTOR MECHANISM

K_s = FUNCTION OF SURFACE HEAT LOSS MECHANISM

the other central. The skin thermoreceptors are a part of the peripheral system consisting of the skin receptor, thalamocortical relays, and the peripheral musculature. Information on skin temperature is received and processed by this system and, when necessary, is translated into highly coordinated motor activity. The central thermoregulatory system includes the hypothalamic thermoreceptor and its effectors, which are primarily the sweat glands and the cutaneous blood vessels.

The heat loads handled by the sweat glands and the dilation and constriction of the cutaneous blood vessels are impressive. Water can be excreted at a rate of more than two gallons a day, dissipating four times the heat generated metabolically. Heat conductance to the skin can increase over 7 times the standard level, an increase equivalent to more than two tons of blood circulated through the skin in 24 hours. However, it is mainly through control of the temperature and humidity of the external environment, and the use of various types of clothing, that man is enabled to survive and even proliferate in hostile environments. The two independent systems of heat regulation, working together, provide the basis for both the extremely wide range of environmental temperatures than can be made conducive to man, and for the precision of internal thermoregulation.

5.7 Infrared Detectors in Snakes

This subject has been covered extensively in the technical, semi-technical and popular literature, and, therefore, need be reviewed only cursorily here. The definitive work in the area may be found in Bullock and Diecke (Ref. 5-64) and Bullock and Fox (Ref. 5-65).

The essential features are as follows. As shown in Figure 5-9A, the pit containing the sensor element is located beneath and just anterior to the eye. A cross-sectional diagram of a pit appears in Figure 5-9B. The pit membrane, which is the sensor element, is 3-4 mm² in area. The membrane is about 15 μ thick, and contains 500-1500 nerve fiber endings per square millimeter. These nerve endings lie 5-7 μ below the surface facing outward. The responses of these nerves to stimulation can be recorded from the centripetally-conducting nerve, or at the membrane. The snakes were immobilized by the intramuscular injection of curare. Under these experimental conditions, it was found that the peak responses occurred upon stimulation with

FIG. 5-9A
Infrared Pit Sensor of the Viper

Head of 'Crotalus' cut away to show the facial pit and its sensory membrane. The membrane is shown as wrinkled to suggest its texture and thinness and to emphasize the air chamber behind it.

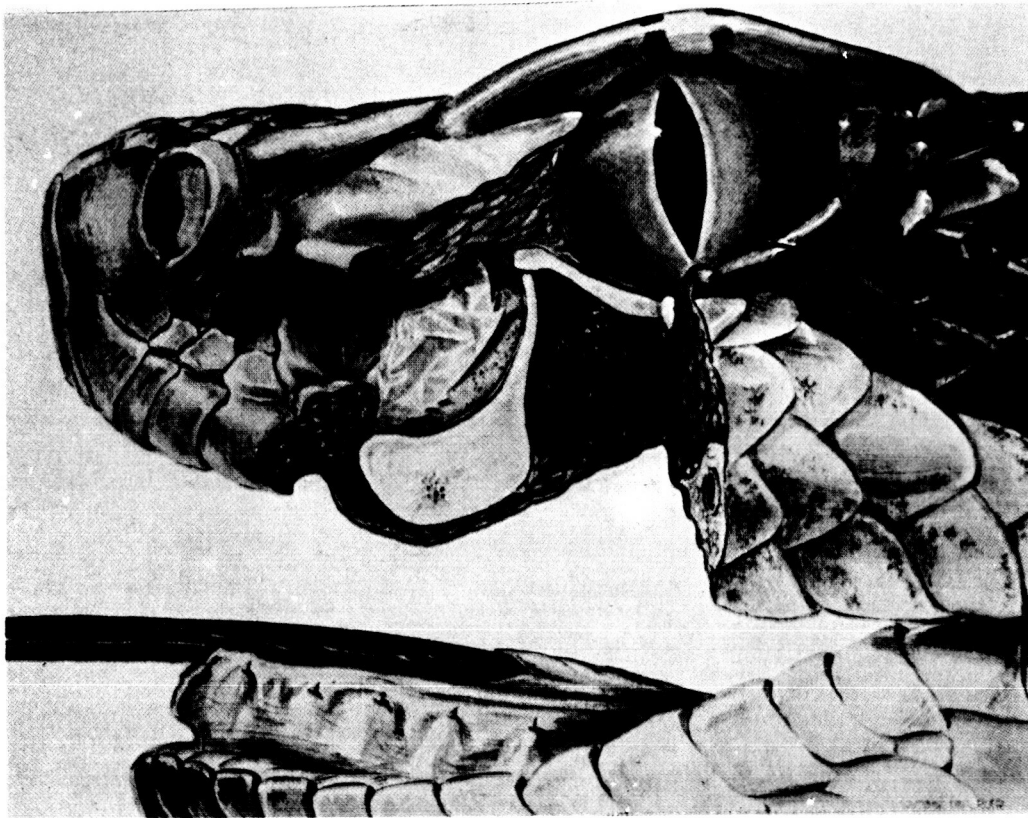
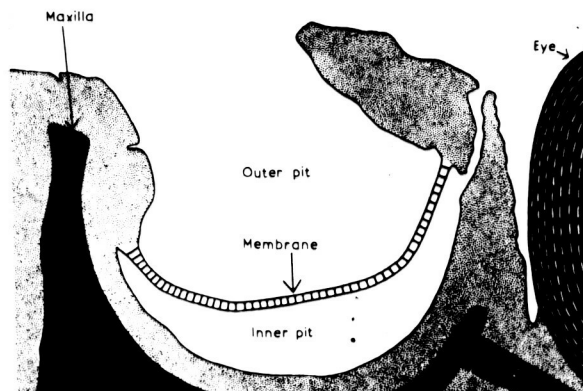


FIG. 5-9B
Cross Sectional Diagram of a Pit

The pit membrane is about 15 microns thick, well vascularized and innervated.



infrared radiations in the 1.5 - 4.0 μ wavelength range. No responses were recorded below 1 μ , nor above 10 μ , and responses to the 5-7 μ range were unreliable. The latency (the time lag between the presentation of the stimulus and the first neural response) varied from 10 to 50 msec., the shorter times being observed when higher-intensity stimulation was used. The maximum frequency of firing of a single nerve cell was 180/sec. The intensity range was usually less than 500 times the threshold level.

Unlike the temperature receptors in the tongue of the cat, which fired at rates that were somewhat proportional to the temperature ("tonic") (see Fig. 5-5), the pit receptors adapted quickly ("phasic"), that is, fired rapidly at the onset of stimulation but soon returned to a prestimulation random firing rate even though the stimulus remained on. A "tonic" receptor may be compared to a direct-coupled amplifier, and the "phasic" to a condenser-coupled amplifier, in terms of the output to a DC input. The pit sensor therefore transmits information on the change of radiant flux, rather than the ambient conditions. This is an interesting property of many biological sensory systems, such as touch and vision. (In the case of vision, even though the receptors are phasic, vision is maintained because of the small, spontaneous tremors of the eyeball; see, for example, Pritchard, Heron, and Hebb, 1960). The advantages of such a system to a predator like the viper is that simply by holding its head still, it is no longer stimulated by steady-state objects within its operative range. A warm object moving across its cone of reception would therefore produce a signal against an essentially silent background. Since the steady-state background is "noise" as far as the snake is concerned, by using phasic receptors, the noise level is drastically reduced. By this simple technique the snake is provided with an infrared moving-target detector with a great signal-to-noise ratio, using components which are inherently noisy.

A simple infrared moving-target detector can be designed using these principles (Fig. 5-10). Since the amplifiers have short time-constants, steady-state or slow-changing ambient conditions are blocked. However, a warmer object that moved across the field of view would produce successive transient outputs. Although such a device could conceivably find application as a sensor in a tracking or missile homing system, or to detect intruders in a restricted zone, problems may arise in achieving the degree of resolution required while maintaining complexity within reasonable limits.

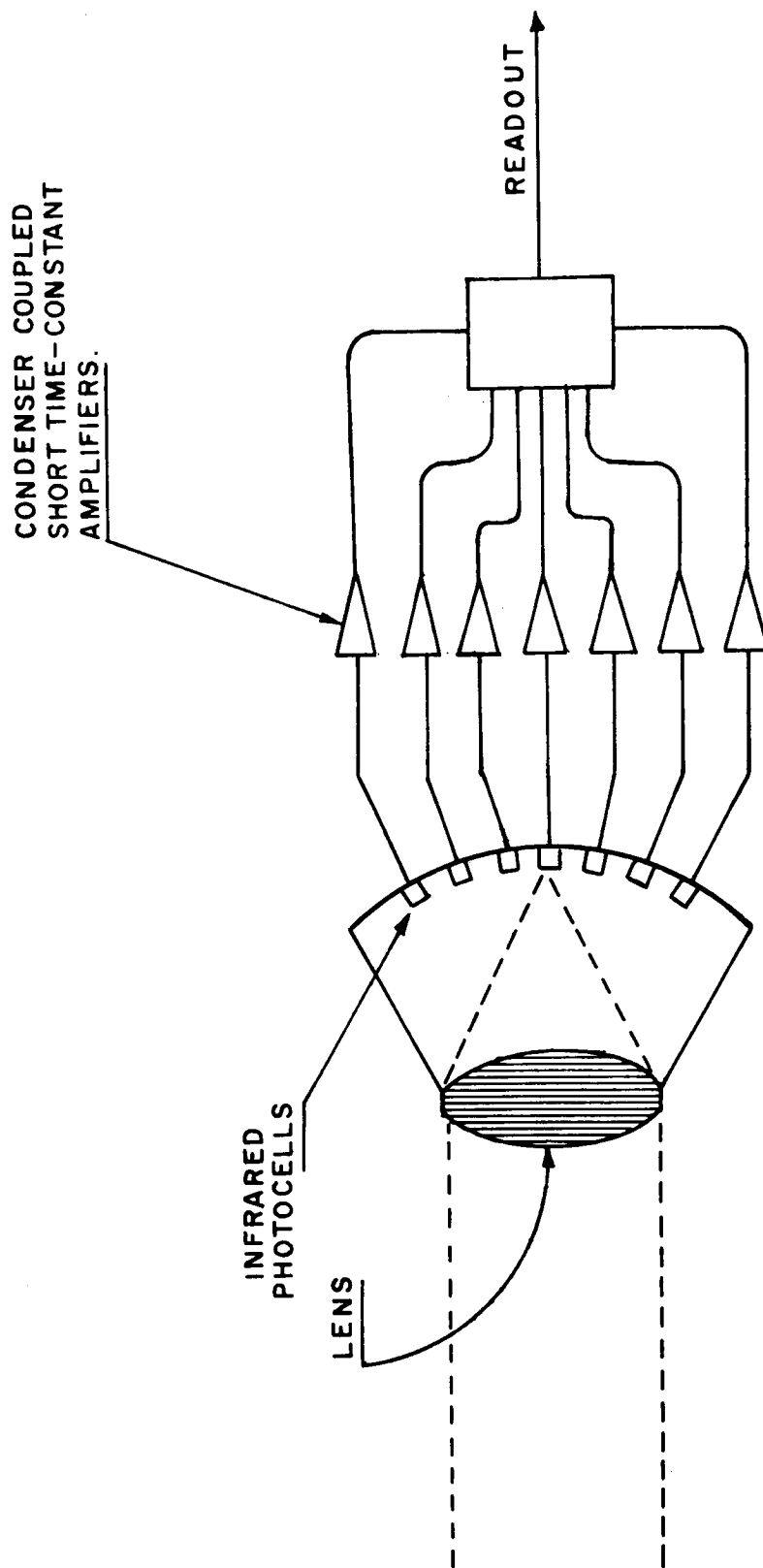


FIG. 5-10

SCHEMATIC DIAGRAM OF AN INFRARED MOVING TARGET INDICATOR

The cone of perception of the snake pit sensors is diagrammed in Figure 5-11. These have apparently been determined in the curarized snake, which is unable to make normal muscular movements (Ref. 5-64). There is a possibility that under normal conditions a snake can vary the extent and direction of reception. Referring to Figure 5-9B, note that there is a muscular structure in place at the mouth of the pit. This structure may function as a control valve, operating in such a manner as to provide the snake with a "search" and a "track" mode. This mechanism would certainly enhance the survival capabilities of the snake. If the pit sensor operates in this manner, biologists will undoubtedly point out that once again nature had anticipated engineers by several hundred million years. These observations show that the study of animals can produce new engineering principles. Accelerated efforts can uncover advanced concepts for engineering applications.

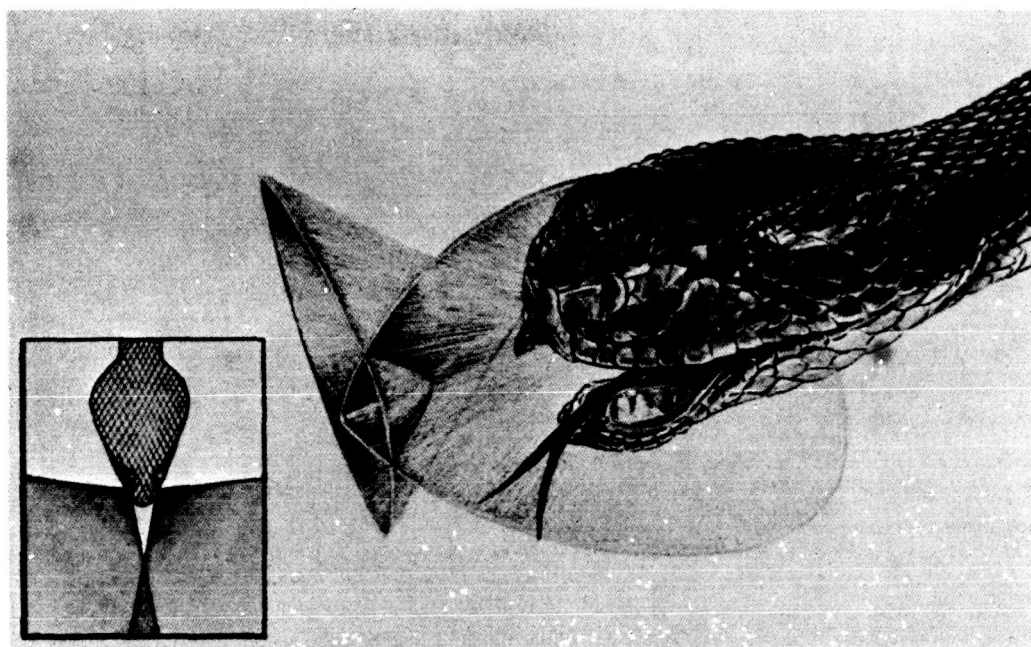
Much has been made of the sensitivity of the snake pit sensor, which can detect temperature changes of $.001 - .003^{\circ}\text{C}$, depending of the method of measurement, the lower value being calculated from the heat flux and the higher value being observed when the sensor is warmed by running water. However, in man a temperature change of as little as $.003^{\circ}\text{C}$ in 3 seconds can be detected at the forehead (Ref. 5-36 and 5-66).

Principally because of the high energy flux that is needed to produce a threshold response in the snake, the pit sensor is most likely a temperature sensor (Ref. 5-64). It has been suggested that the pit may operate like a Golay radiometer, where the applied radiation warms the interior chamber (see Fig. 5-9B), the resulting increase in volume being detected by the membrane (the membrane then functioning as a stretch receptor). This seems unlikely for a number of reasons, the most cogent being that normal functioning is not disturbed by puncturing the membrane.

5.8 The Ampullae of Lorenzini

The functions of a specilized type of receptor found in large numbers in the head region of sharks and rays, the ampullae of Lorenzini, are thought by Sand (Ref. 5-67) to be thermoreception. The earlier history of the study of the ampullae can be found in Dotterweich (Ref. 5-68). In general, the ampullae are canals filled with jelly-like substance which does not give a specific mucus reaction. The canals can be of considerable length and can

FIG. 5-II
Cone of Reception of the Pit Sensor



reach up to 4 mm in diameter. Several types of cells are found in the base of the ampullae but their functions have not yet been isolated. Some behavioral data suggest that the ampullae serve as absolute pressure sensors, a decrease in the viscosity of the jelly-like substance being correlated with an increase in ambient pressure. However, Sand (Ref. 5-67), recording from the afferent nerves of the ampullae, found a massive "spontaneous" discharge that could not be influenced by touch, stroking, or pressure. On the other hand, a highly sensitive response occurred to a warm or cool solution dropped on the ampullae (see also Ref. 5-69). Paradoxically, the discharge rate increased to cooling and decreased to warming (Fig. 5-12). Note that the responses are tonic, not phasic. If these ampullae are primary temperature sensors, the peculiar design, in which the nerve elements are covered with the jelly-like substance, is hard to understand. That they are sensitive to local changes of water pressure near the surface of the body has been shown recently by Lowenstein (1960) and by Murray (1957). The suggestion, then, is that the lateral line system senses rapid changes in water pressure or flow, and the ampullae react to slower and longer lasting pressure effects, as Dotterweich originally suggested. However, the details of the mechanism by which the pressure produces the nerve discharge is not known.

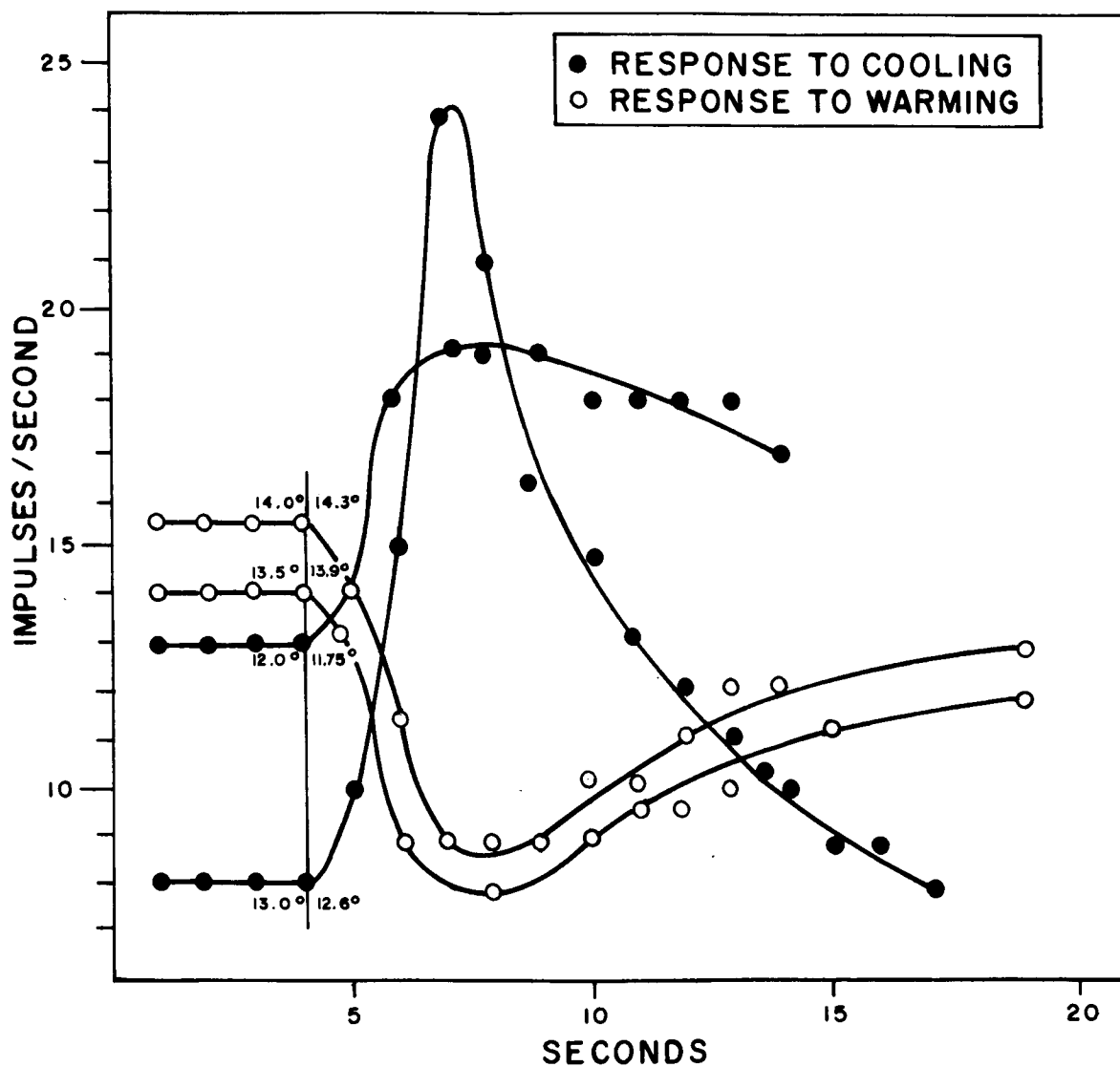
5.9 Conclusion

This review of the current state of biological knowledge of temperature sensors in animals has revealed some limitations of existing data for engineering purposes. Although some comments may appear to have been overly critical, the guiding principle of this program is to take a good, hard look at data compiled by the biologists. In a way this is an unfair evaluation of their efforts, since their aim as biologists is not to collect data for the express purpose of determining possible engineering applications. On the other hand, such an aim may give added guidance, coherence and information to biological research.

A factor that limits the usefulness of research or data in this area is the relatively narrow temperature range over which biological temperature sensors operate, which is well within the 0-100° C. Physical temperature sensors are available that cover the spectrum from absolute zero to several million degrees. Moreover, physical sensors can be designed with sensitivities

FIG. 5-12

RESPONSES OF SINGLE UNITS OF A MANDIBULAR
AMPULLA OF LORENZINI TO TEMPERATURE CHANGES
OF LESS THAN 0.5°C .



at the quantal limit, which is several orders of magnitude greater than the most sensitive biological thermoreceptor. However, a detailed study may lead to new concepts of construction of thermal sensors.

On the other hand, the data seem to show conclusively that all biological thermoreceptors are temperature sensors. This means that at the potential generator site of the receptor, a small change in the temperature of the surrounding fluid medium is sufficient to trigger the neural action potential, or in the case of the hypothalamic thermoreceptors, the change of electrical field potential. This returns us to one of the central problems of biology, viz. the mechanism of neural excitation and conduction. The solution to this question will have important implications for the theory and practice of information transmission. An understanding of central thermoregulation may then indicate how this information is processed.

REFERENCES - SECTION 5

- 5-1 Boring, E. G., Sensation and Perception in the History of Experimental Psychology. D. Appleton-Century Co., New York, (1942)
- 5-2 Blix, M. Uppsala Läkaref. Förh. 18, 87, (1882)
- 5-3 Beetz, F. Arch. Gynäk. 162, 106, (1936)
- 5-4 Hauer, P. Ztschr. Biol. 85, 265, (1926)
- 5-5 Hirsch, L. and Schriever, H. Ztschr. Biol. 89, 1, (1930)
- 5-6 Schriever, H. and Strughold, H. Ztschr. Biol. 84, 193, (1926)
- 5-7 Rein, H. Ztschr. Biol. 82, 189, (1925)
- 5-8 Speiser, M. Arch. Gynäk. 146, 137, (1931)
- 5-9 Strughold, H. Verhandl. Phys.-Med. Gesellsch. Würzb. 51, 31, (1926)
- 5-10 Strughold, H. and Karbe, M. Ztschr. Biol. 83, 189, (1925)
- 5-11 Strughold, H. and Porz, R. Ztschr. Biol. 91, 563, (1931)
- 5-12 Hensel, H. Ergebn. Physiol. 47, 165, (1952)
- 5-13 Goldscheider, A. Arch. Ges. Physiol. 39, 96, (1886)
- 5-14 von Skramlik, E. Arch Psychol. 4, 244, (1937)
- 5-15 Bazett, H. C., McGlone, B., Williams, R. G., and Lufkin, H. M. A.M.A. Arch. Neurol. and Psychiat. 27, 489, (1932)
- 5-16 von Frey, M. Ber. Sächs. Gesellsch. Wiss. 47, 166, (1895)
- 5-17 Alrutz, S. Skandinav. Arch. Physiol. 7, 321, (1897)
- 5-18 Ebbecke, U. Arch. Ges. Physiol. 169, 395, (1917)
- 5-19 Endres, G. Ztschr. Biol. 89, 536, (1930)
- 5-20 Thunberg, T. Skandinav. Arch. Physiol. 12, 394, (1902)
- 5-21 Hensel, H. and Zotterman, Y. J. Neurophysiol. 14, 377, (1951)
- 5-22 Cohen, M. J., Landgren, S., Ström, L., and Zotterman, Y. Acta. Physiol. Scandinav., Suppl. 135, (1957)
- 5-23 Landgren, S. Acta Physiol. Scandinav. 40, 202, (1957)
- 5-24 Hensel, H. and Zotterman, Y. Acta Physiol Scandinav. 22, 96, (1951)
- 5-25 Hensel, H. and Zotterman, Y. Acta Physiol Scandinav. 23, 291, (1951)
- 5-26 Hensel, H. Arch. Ges. Physiol. 256, 195, (1952)
- 5-27 Zotterman, Y. Skandinav. Arch. Physiol. 75, 106, (1936)

- 5-28 Dodt, E. and Zotterman, Y. Acta. Physiol. Scandinav. 26, 258, (1952)
- 5-29 Hensel, H. and Zotterman, Y. J. Physiol. 115, 16, (1951)
- 5-30 Lele, P. P., Weddell, G., and Williams, C. M. J. Physiol. 126, 206, (1954)
- 5-31 Bazett, H. C. and McGlone, B. A.M.A. Arch. Neurol. and Psychiat. 27, 1031, (1932)
- 5-32 Hensel, H. Arch. Ges. Physiol. 252, 165, (1950)
- 5-33 Cheng, J. Acta. Scientiarum Naturalium Univ. Pekingensis, 1, 95, (1959)
- 5-34 Dodt, E. Arch. Ges. Physiol. 263, 188, (1956)
- 5-35 Dodt, E. and Walther, J. Arch. Ges. Physiol. 256, 355, (1957)
- 5-36 Hardy, J. D. and Oppel, W. J. Clin. Invest. 17, 771, (1938)
- 5-37 Tomita, T. J. Neurophysiol. 21, 419, (1958)
- 5-38 Barbour, H. G. A. Res. Nerv. and Ment. Dis., Proc. 20, 449, (1940)
- 5-39 Bazett, H. C. In: Physiology of Heat Regulation and the Science of Clothing. Newburgh, L. H., ed., Saunders, Philadelphia, (1949) p. 109.
- 5-40 Beattie, J. In: The Hypothalamus. LeGros Clark, W. E., Beattie, J., Riddoch, G., and Dott, N. M., ed., Oliver, London, (1938) p. 69.
- 5-41 Bronk, D. W., Pitts, R. F., and Larrabee, M. G. A. Res. Nerv. and Ment. Dis., Proc. 20, 323, (1940)
- 5-42 Davison, C. A. Res. Nerv. and Ment. Dis., Proc. 20, 774, (1940)
- 5-43 Dubois, E. F. Fever and the Regulation of Body Temperature. Thomas, Springfield, (1948)
- 5-44 Folkow, B. Physiol. Rev. 35, 629, (1955)
- 5-45 Hensel, H. Ergebn. Physiol. 43, 166, (1952)
- 5-46 Hess, W. R. Helvet. Physiol. et Pharmacol. Acta., Suppl. IV, (1947)
- 5-47 Newburgh, L. A., ed. Physiology of Heat Regulation and Science of Clothing. Saunders, Philadelphia, (1949)
- 5-48 Ranson, S. W. A. Res. Nerv. and Ment. Dis., Proc. 20, 342, (1940)

- 5-49 Ranson, S. W. and Magoun, H. W. *Ergebn. Physiol.* 41, 56, (1939)
- 5-50 Walker, E. A. A. *Res. Nerv. and Ment. Dis., Proc.* 20, 400, (1940)
- 5-51 Finley, H. K. A. *Res. Nerv. and Ment. Dis., Proc.* 20, 286, (1940)
- 5-52 Magoun, H. W., Harrison, F., Brobeck, J. R., and Ranson, S. W. *J. Neurophysiol.* 1, 101, (1938)
- 5-53 Beaton, L. E., McKinley, W. A., Berry, C. M., and Ranson, S. W. *J. Neurophysiol.* 4, 478, (1941)
- 5-54 Eliasson, S. and Ström, G. *Acta. Physiol. Scandinav.* 20, Suppl. 70, 113, (1950)
- 5-55 Folkow, B., Ström, G., and Uvnäs, B. *Acta. Physiol. Scandinav.* 17, 47, (1950)
- 5-56 Hemingway, A., Rasmussen, T., Wikoff, H., and Rasmussen, A. T. *J. Neurophysiol.* 3, 329, (1940)
- 5-57 Ström, G. *Acta. Physiol. Scandinav.* 20, Suppl. 70, 47, (1950)
- 5-58 Ström, G. *Acta. Physiol. Scandinav.* 20, Suppl. 70, 8, (1950)
- 5-59 Ström, G. *Acta. Physiol. Scandinav.* 21, 271, (1950)
- 5-60 von Euler, C. J. *Cell. and Comp. Physiol.* 36, 333, (1950)
- 5-61 von Euler, C. *Acta. Neuro. Veg.* 3, 113, (1951)
- 5-62 von Euler, C. and Söderberg, U. *J. Physiol.* 118, 555, (1952)
- 5-63 Benzinger, T. H. *Proc. Lab. Acad. Sci. (Wash.)*. 45, 645, (1956)
- 5-64 Bullock, T. H. and Diecke, T. P. S. *J. Physiol.* 134, 47, (1956)
- 5-65 Bullock, T. and Fox, W. *Quart. Jour. Microscop. Sci.* 98, 219, (1957)
- 5-66 Hardy, J. D. and Oppel, W. *J. Clin. Invest.* 16, 533, (1937)
- 5-67 Sand, A. *Proc. Roy Soc., London. ser. B* 125, 524, (1938)
- 5-68 Dotterweich, H. *Zool. Jahrb. Abt. 3.* 50, 347, (1932)
- 5-69 Hensel, H. *Arch. Ges. Physiol.* 263, 48, (1956)
- 5-70 Hardy, J. D. *Harvey Lectures*, 49, 242, (1954)

6. The Neural Aspect of Sensory Function and It's Application

6.1 Introduction

A realistic program to apply biosensor characteristics to development of new instrumentation concepts must consider the "modus operandi" of the nervous system as a whole, the transmission characteristics of neural cells and the detailed physical and chemical mechanisms of interaction of the electrical neural signal with biological structures. In the normal development of an engineering system to utilize a particular sensor to instrument a specific problem, the interaction between the requirements of the sensor and the rest of the system determine the design parameters of both. In the present approach, the designer seeking to utilize biological principles is presented with a "fait accompli"; a very efficient and economical system performing its required functions, field tested and modified through evolutionary processes. Thus the examination of the sensor is most fully productive when considering its integrated function.

Ultimately, any sensor system, physical or biological must collect input information and transmit it with minimum attenuation and distortion; retaining as much information content as possible in the process. In terms of the neural process, this means that both the "transmission line" itself, the axon, must be considered as well as the various matching phenomena between neurons and the optical, mechanical or other components of the sensor responsive to the environment. The unusual nature of the transmission process (unusual compared to the systems met with in engineering practice) has been viewed for some time with considerable interest by bionicists. It is a continuous, active transmission system with chemical energy stored along the transmission path. Triggering and pulse shaping are determined by chemical processes. A widely varying group of "neuristor" concepts have developed for ultimate application to information systems.

Detailed electromagnetic studies such as the radiative characteristics and the specific current and voltage relationships during initiation and propagation of the neural impulses can be extremely important in several bioastronautic problems. For example, the visual and aural routes for information input are severely overloaded during critical flight phases such as re-entry. Various schemes to utilize other sensory mechanisms such as

the tactile or olfactory sense have been considered. However, to date, these are only gross information handling systems. A desideratum would be some method of direct input of sensory information into the central nervous system derived from various sensing instruments or communication channels. Similarly, a direct output to external instrumentation would afford new and interesting possibilities for control. Another problem the astronaut may face is maintaining for an extended time his full operational capability in a strong magnetic field, possibly with strong gradients. A magnetic field may be the preferred method of reducing charged particle fluxes for extended space missions. The effect the magnetic field will have on the whole organism will ultimately reduce to the specific problem of the combined action of an electric and magnetic field on the neural currents. A study of the effects these fields on single nerves could reveal the basic nature of the effect and make possible the determination of critical threshold, exposure strength, duration, cumulative effect, etc.

Another aspect of potential application to new methods of instrumentation resides in the mechanism of chemical and electrical interaction with the specific cell membranes. The control of permeability utilizing chemical action and subsequent hydrolysis phenomena to damp the process may furnish a new approach using membranes and electrochemical action for sensors to detect, say, micrometeorites or the presence of viable life forms requiring minimum size, voltage and power. Examination of the physical and chemical principles so successfully employed in nature using biological materials will contribute the guidelines for developing similar systems, perhaps using simpler materials.

A brief discussion of the philosophy motivating this section is appropriate. One school of thought in bionics places strong emphasis on modeling devices as closely as possible on biological organisms. Another school relies on biological mechanisms for providing motivation and for assisting in conceiving simple components for engineering combination into complex system. These two schools of thought are complementary and are both used without apology and interchangeably throughout the discussion.

This discussion of nerves, the nervous system and related devices is in two parts. Section 6-2 is purely scientific discussion with no explicit reference to practical applications. The subject matter to be discussed, is of course, vast and the scientific discussion is, therefore, highly selective. The criterion for selection is relevance to the material of Section 6-3 which treats applications of bionics interest.

6.2 Biological Discussion

The biological discussion begins with the nerve axon. The reason for this is that the axon is relatively simple in structure but nevertheless incorporates many of the complexities of nerve cell operation. The behavior of the axon membrane is of basic importance and this is discussed next. The biological discussion is closed with comments on the general organization of the nervous system.

6.2.1. The Nerve Axon

Descriptive

The nerve axon is the long distance transmission line of the nervous system. It is not surprising, therefore, that the axon has approximately cylindrical symmetry in its structure and this symmetry is important to the experimentalist.

Nerve axons come in a bewildering variety of forms and environments. Most of the experiments discussed below were performed either on vertebrate sciatic nerves or on the giant axon of the squid. In both cases the nerve axons are associated with Schwann cells. In the sciatic nerve, the individual axons may or may not be myelinated. The myelin is formed of layers of Schwann cell membrane and forms a sheath around the axon which is interrupted at sites called Ranvier nodes. This is discussed below. The squid giant axon is not myelinated and, in fact, myelination is rare among invertebrates.

Figure 6-1 is a schematic diagram of the cross section of a nerve fiber consisting of a Schwann cell and four associated axons, (Ref. 6-1). The axons are embedded to varying degrees. The Schwann cell is, however, simply connected, that is, the membrane of the Schwann cell can be traced with one continuous line without intersections. Note that the membranes of the axons and of the Schwann cell are drawn as double layers. When the axon is completely embedded in the Schwann cell it is called a mesaxon. The significance of the double layer appearance of the membrane will be discussed in Section 6.2.2.

A high power micrograph of a nerve fiber showing unmyelinated axons is shown in Figure 6-2, (Ref. 6-1). This was obtained using an electron microscope. The similarity to Figure 6-1 is apparent.

Nerve fibers of the above type are found in both invertebrates and vertebrates. In a very few invertebrates and in all vertebrates an interesting phenomenon can be observed. The membrane of the Schwann cell grows around the axon to form what is called the myelin sheath. This is illustrated schematically in Figure 6-3, (Ref. 6-1). As would be expected in a continuous growth process, the Schwann cell remains simply connected. The cytoplasm, however, is evidently absent in the myelination

FIG. 6-1

Diagram of a typical adult unmyelinated nerve fiber showing a Schwann cell with four associated axons, a-d. A mesaxon m is formed around axon d. A segment of an axon-Schwann membrane is enlarged to the lower right to show the gap between the two unit membranes.

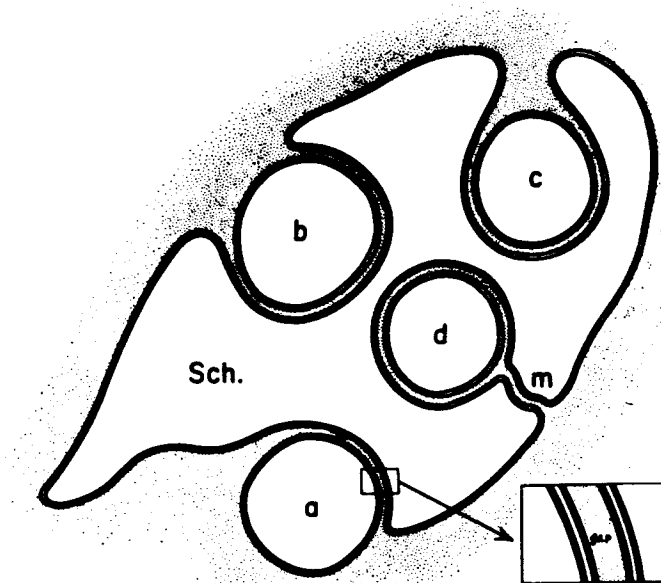


FIG. 6-2

High-power micrograph of an unmyelinated nerve fiber showing a Schwann cell (Sch.) with several associated axons (ax.). Note the unit membranes and the gap g. between the axon and Schwann cell membrane; m. is mesaxon (X152,000).

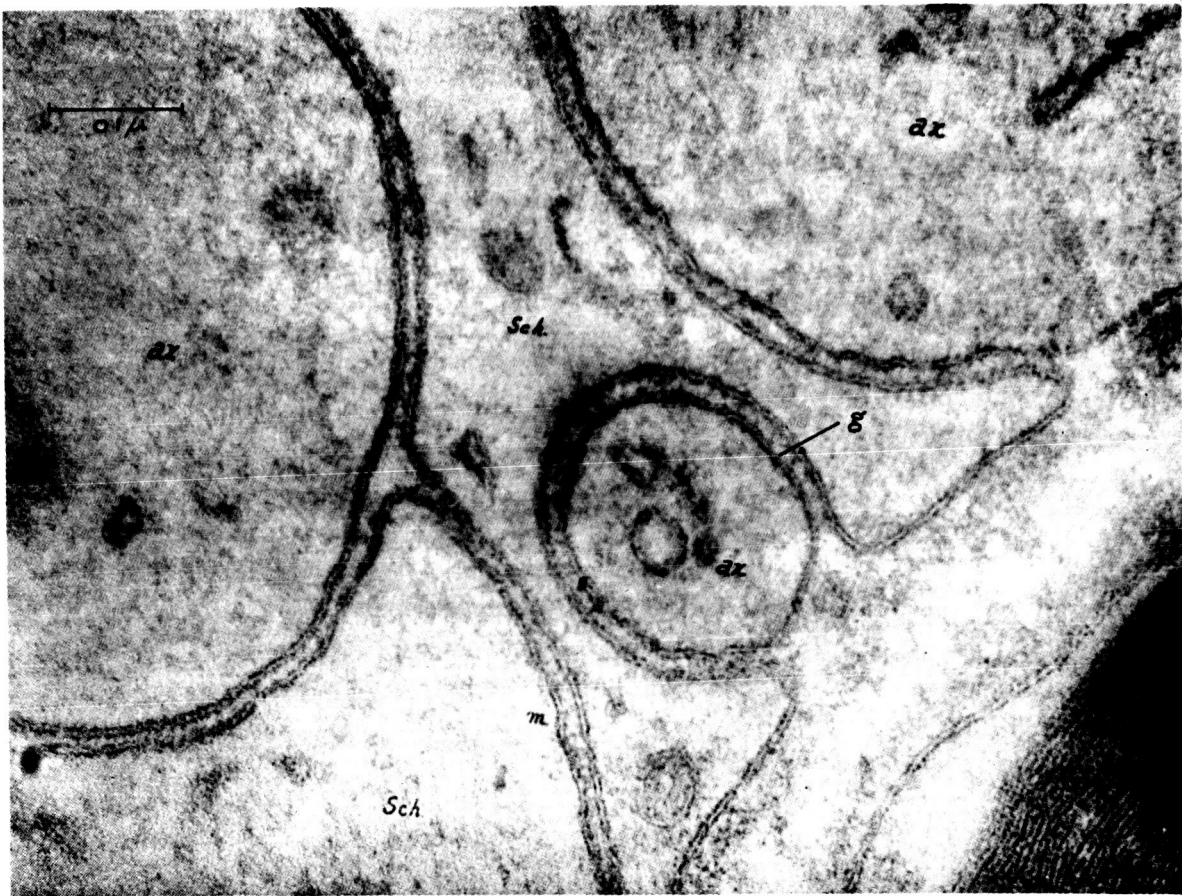
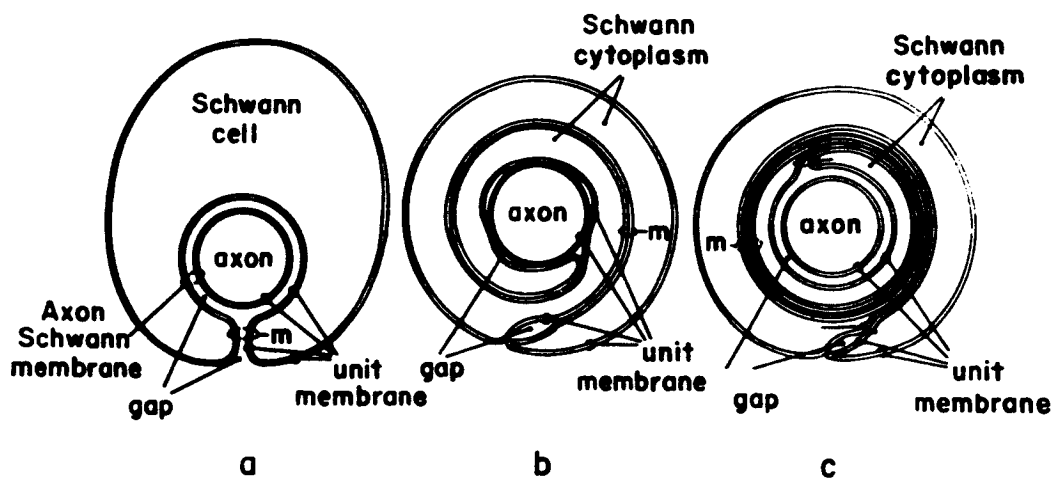


FIG. 6-3

Diagram summarizing the formation of myelin. In (a) a single Schwann cell with a short mesaxon surrounds an axon. This type of fiber is called a protofiber. In (b) the gap between the unit membranes and the mesaxon is obliterated and the mesaxon is elongated in a spiral around the axon. This type of fiber is called an intermediate fiber. In (c) the mesaxon loops have come together along their cytoplasmic surfaces to make the major dense lines of compact myelin. This type of fiber is designated by the term "myelinated."



and the myelin has the appearance of a number of layers of membrane.

Myelination is a growth process which is observable in young vertebrates. Figure 6-4 is a micrograph of a section from a young mouse sciatic nerve, (Ref. 6-1). All stages of myelination can be seen.

The myelination of a nerve axon is interrupted as shown in Figure 6-5, (Ref. 6-2). The interruptions are called Ranvier nodes. The node actually has a complicated structure as shown in Figure 6-6, (Ref. 6-3). Most of the structures shown are not of immediate interest, but it will be noted that the axon membrane is continuous through the Ranvier node and that it is possible to go from the exterior of the system to the axon membrane without passing through the myelin sheath.

The structure of the Ranvier node varies systematically with fiber diameter as shown in Figure 6-7.

In subsequent sections certain experiments dealing with the behavior of nerve axons will be discussed. The difference between myelinated and unmyelinated fibers presumably has functional significance. It will become apparent, however, that the basic mechanism of nerve impulse transmission is almost certainly the same in both types of axon.

The cylindrical symmetry of the nerve axon has been referred to as being experimentally convenient. The geometry is not perfect, especially for myelinated axons, but the geometry is close enough to having cylindrical symmetry for experimental convenience.

General Experiments and Theory

Historically, early experiments on nerve axons investigated the phenomenon of electrical stimulation of nerve impulses. This is a historical accident resulting from the ease of placing external electrodes. Since electrical excitation involves a complex set of more or less parallel phenomena it is more convenient to discuss first experiments dealing only with individual simple phenonema within the set.

The first question to be raised is: how does one observe an axon impulse? The most obvious method is that the existence of the impulse can be detected by the impulse performing its biological function; i. e., causing a muscle to twitch or registering an effect in the brain. Many experiments have been performed with the frog sciatic nerve-

FIG. 6-4

A low-power survey micrograph of a transection of a young mouse sciatic nerve fiber showing several stages in the formation of myelin. The earliest stages are represented by protofibers (P.). The later stages in which the mesaxon is elongated in a spiral about the axon are designated I, and the later stages in which myelin has been formed are labeled myl. (X9, 500).

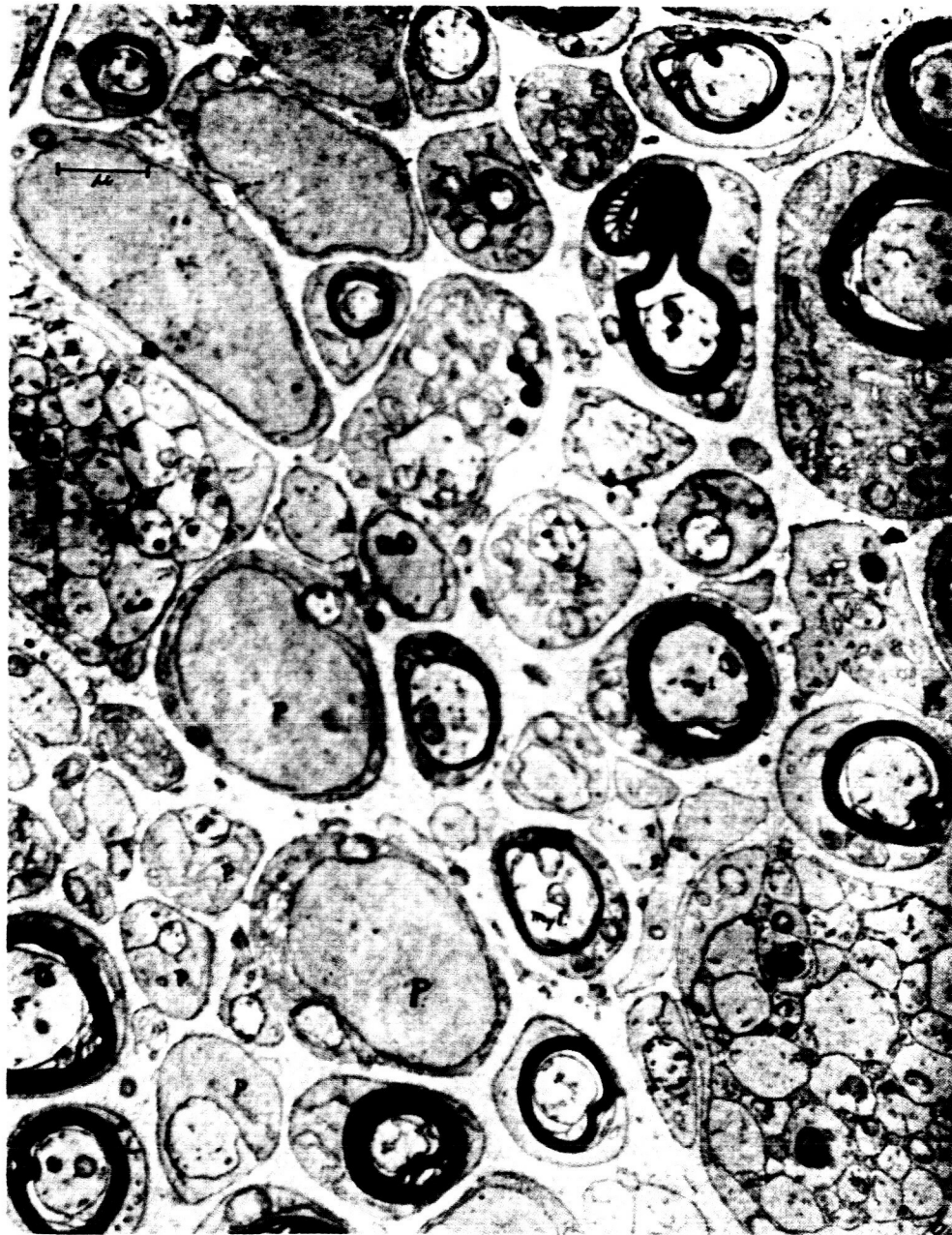
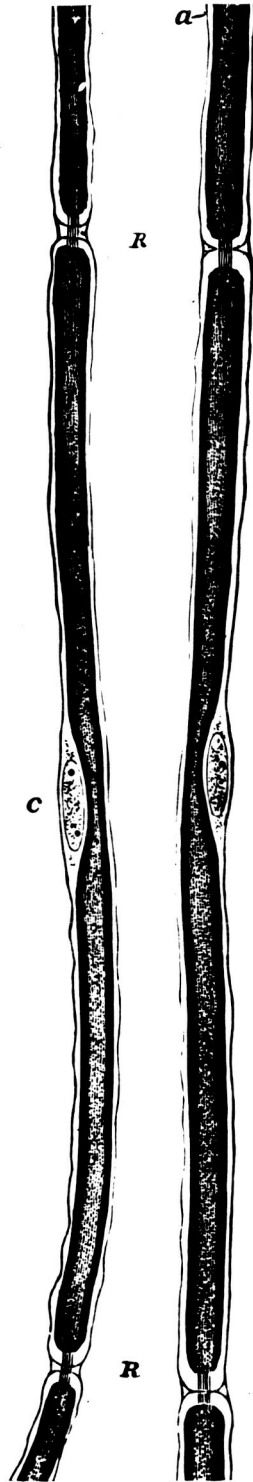


FIG. 6-5

Diagram of medullated nerve fibers stained with osmic acid. X425.
R. Nodes of Ranvier. a. Neurolemma. c. Nucleus.



6-10

FIG. 6-6

Diagram summarizing the various structures observed at nodes. The nodal axoplasm contains moderately abundant tubular components of the endoplasmic reticulum (e.r.), mitochondria (M)-multivesicular bodies (m.v.), and large, very dense, presumably lipid bodies (L). The nodal axoplasm bulges slightly at the node. The juxtaterminal myelin region is as described in Text Fig. 13. Sometimes, however, the mesaxons (m) are somewhat complicated in appearance by the retraction of some of the terminal rings of Schwann cytoplasm (Sch.), as indicated to the lower right. The axon-Schwann membrane gap is closed over the juxtaterminal myelinated region. At the unmyelinated nodal region the gaps are present and measure about 100-150 Å across. This region is covered by numerous fingerlike processes (pr.) extended by the two Schwann cells meeting at the node.

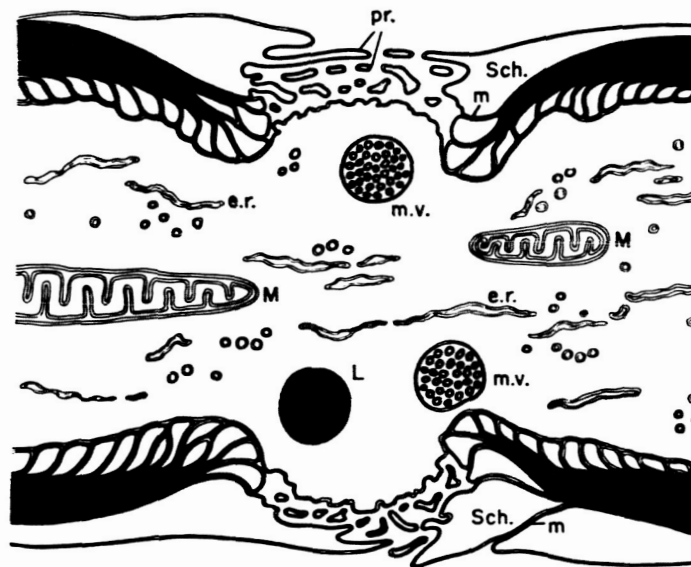
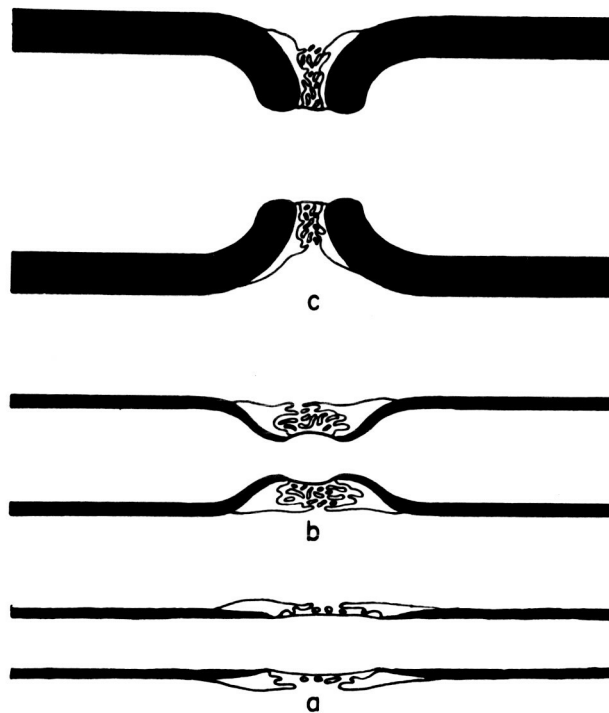


FIG. 6.7

Diagram illustrating the arrangement of the collar of nodal processes of Schwann cells about nodes of Ranvier in fibers of different sizes. In the smallest fibers the collar is relatively thin as in (a) and (b) but it is spread over a relatively greater length. In larger fibers, the thickness of the collar is much greater, but its extent along the axon is reduced (c).



gastrocnemous muscle preparation. This method is still not without its value. Modern instruments, first the string galvanometer and more recently the electron beam oscilloscope, permit impulse detection by electrical methods. More recently, Gengerelli, et al., (Ref. 6-4) observed magnetic fields in the vicinity of a nerve impulse. This may prove a valuable method in the future.

A typical experimental setup is shown in Figure 6-8, (Ref. 6-5). The nerve is excited. This excitation may be accomplished by electrical means, by deforming the nerve, by temperature change or by action of a sensor at the end of the nerve. The impulse is transmitted along the nerve. As the impulse passes, the first electrode becomes negative to the second electrode. Then the second electrode becomes negative relative to the first. The oscilloscope trace thus produced is shown in Figure 6-9. If the electrodes are close enough the diphasic pulse of Figure 6-8 is seen. If the nerve is crushed or narcotized under one electrode, one obtains a monophasic oscilloscope trace as shown in Figure 6-9. These potential changes are referred to as the "action potential." Their significance is discussed below.

Under "Descriptive" above, the axon was described as a membrane surrounding a core of protoplasm. The question arises of the relationship among the action potential, the membrane, and the core. It would be desirable to remove the protoplasm, replace it with appropriate ionic solutions and observe the resulting effects on the action potential. This experiment has only recently been performed by Baker, Hodgkin and Shaw, (Ref. 6-6). Although additional results will be discussed below, the following results quoted from Reference 6-6 are of immediate interest.

- "1. Electron micrographs indicate that the axon membrane and Schwann cell layer were still present in extruded and perfused axons. They also show that about 95% of the axoplasm was removed by extrusion and subsequent perfusion.
2. Giant nerve fibers of Loligo, from which the axoplasm had been extruded, conducted impulses if the sheath was refilled either with natural axoplasm or with isotonic solutions containing salts of potassium.
3. The extruded 'sheath' could also conduct impulses, but in that case the conduction velocity was low and the externally recorded action potential was small.
4. Axons filled with isotonic solutions of potassium

FIG. 6-8

THE PASSAGE OF A NERVE IMPULSE IN A NON-CONDUCTING MEDIUM

The resultant recording is obtained as the wave of activity passes consecutively under two electrodes both of which are on the nerve. The record obtained when the two electrodes are close together, as shown at c, gives the appearance of a diphasic wave. It should be noted that in the upper two diagrams the abscissa represents length in space, and in the lower two it represents passage of time.

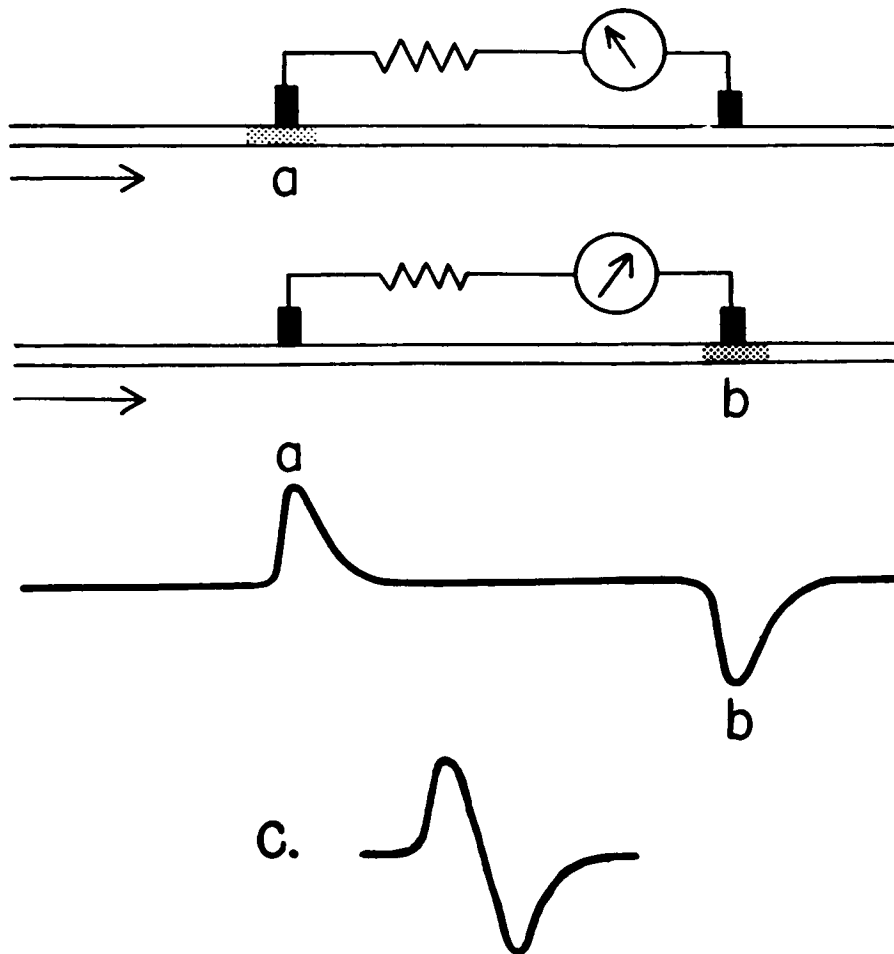
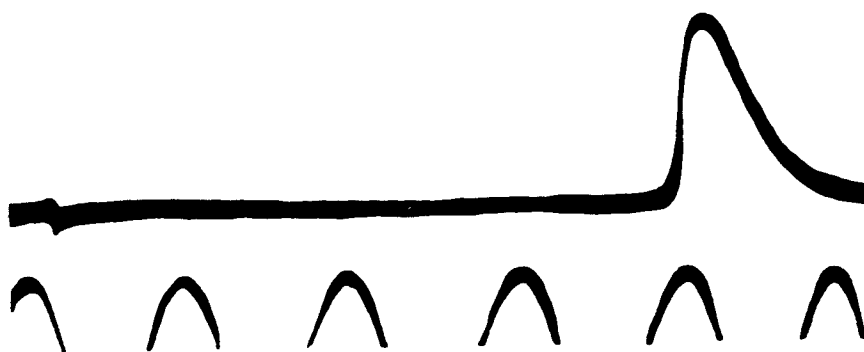


FIG. 6-9

Monophasic spike potential of single unmyelinated axon of crab nerve in oil



isethronate or sulphate remained excitable for up to 5 hrs. and would conduct 3×10^5 impulses."

It is safe to conclude that action potential phenomena are determined primarily by the properties of the membrane as long as the axon contains an isotonic solution of a potassium salt. This tends to indicate that the protoplasm, as such, is involved in long term life maintaining functions* and has no direct significance for action potential phenomena.

Given the above results the question naturally arises, what are the differences between the interior and the exterior of the nerve in the resting state and during passage of the nerve impulse? The most striking differences are in electric potential and in ionic concentration.

It is possible to measure directly this potential difference in the giant axon of the squid (Loligo) and it has been found to be approximately 70 mv. with the interior negative relative to the exterior. This is not appreciably changed when the protoplasm is replaced by appropriate isotonic potassium solutions.

The most striking chemical differences are that potassium ions predominate in the interior of the nerve and sodium ions predominate exterior to the nerve. One would expect to measure a potential difference because of these differences in ionic concentration. Other ions, including calcium and chloride, are also present and seem to influence the conduction process. This is not the same as asserting that these ions actually play a role in the conduction process. Influence has been experimentally demonstrated but an active, significant role has not. In the experiments of Baker, et al., (Ref. 6-6) mentioned above, a large variety of solutions were tried. As long as the sodium and potassium ions were present in appropriate concentrations the axon effectively conducted impulses unless some obviously harmful ionic species were present.

If there is a potential difference between the interior

* Appropriate solutions passed through the cylindrical sheath might permit maintaining the sheath "alive" for a long period of time. This could have instrument application significance as well as fundamental scientific importance.

and exterior of the nerve then there must be a sheath or membrane which has a measurable capacitance and resistance. Using the squid axon, the capacitance has been determined to be about $1 \mu\text{F}/\text{cm}^2$ and the resistance to be $1000 \text{ ohms}/\text{cm}^2$. Similar figures have been obtained for other nerve axons and actually for cell membranes in general. A word of caution is, however, necessary. Quotations of resistance and capacitance must be interpreted in a strictly operational sense as the result of a particular type of experiment. The concepts of resistance and capacitance are macroscopically defined and may not really apply on the molecular and submolecular level.

The first illustration of the caution with which the concept of resistance must be used is that the membrane is selective with regard to ionic passage during the resting state. There is significant potassium permeability but the membrane is quite impervious to sodium.

What happens when a nerve impulse passes?

Appropriate experimental arrangements show that measured membrane capacitance remains practically constant but that the measured membrane resistance decreases at the site of the action potential. The resistance decrease must, however, be reinterpreted as increases in ionic permeability because the permeability changes are selective between sodium and potassium ions. In particular, at the onset of the action potential, the membrane becomes very permeable to sodium ions and somewhat more permeable than in the resting state to potassium ions. The increase in sodium permeability predominates in the initial stage so the inward flow of sodium is most striking. Later in the time history of the action potential, the outward flow of potassium is the dominant ionic flow.

The question arises, what triggers the permeability changes? Experiments of Tasaki (Ref. 6-7), and of Cole (Ref. 6-8) indicate that a relatively small reduction in the potential difference across the membrane initiates a process which reduces the permeability sufficiently for development of the action potential. The permeability reduction results, of course, in an even more rapid decrease in potential difference: the phenomenon is very non-linear.

It is clear now how the action potential propagates. The ionic flows at the site of the action potential cause a reduction in the

potential difference across the membrane at an adjacent site which triggers the permeability changes at the adjacent site and, therefore, the action potential.

It is apparent that some mechanism must restore the resting state ionic concentration after the passage of the action potential. This mechanism must be associated with the membrane or sheath since it operates when the protoplasm is replaced by simple ionic solutions.

The processes discussed above involve energy and one would expect temperature changes resulting from heat evolution and/or absorption should be observable. Since the experimental techniques are difficult, only recently have reasonably satisfactory observations of these temperature changes been obtained (Ref. 6-9). The heat evolution and absorption can be placed in four categories as follows:

1. Resting heat evolution. Beresina, (quoted in Ref. 6-9) has reported that frog sciatic nerve in the resting state releases heat at the rate of 4.14×10^{-3} cal/g. of the nerve per minute at 20°C. Feng, (quoted in Ref. 6-9) has argued that this correlates with observed oxygen consumption.
2. Initial heat evolution. In the initial phase of the action potential heat is evolved.
3. Initial heat absorption. After heat evolution, a heat absorption phase is observed. Slightly less heat is absorbed than is evolved during heat evolution. The time scale of the initial heat evolution and absorption scale is similar to the time of the action potential, but time resolution of the experiments is inadequate to establish exact correlations.
4. Long term heat evolution. This is observed on a time scale about 4×10^3 times longer than that of the initial heat evolution and absorption.

The temperature changes for the last three categories were observed in crab nerve.

It seems reasonable to associate the initial heat evolution and absorption directly with the action potential and the resting potential, and the long term heat evolution with indirect processes not immediately involved in the action potential. A fascinating, but very difficult experiment

to perform would be to observe these heat and temperature effects in axons in which the protoplasm is replaced with ionic solutions. The resting and long term heat evolution (categories 1 and 4) possibly would not occur.

The above discussion deals with logically simple experiments in the sense that each experiment is concerned with only one phenomenon occurring during passage of the action potential. Whenever one of these single phenomena experiments have been performed for, say, both the squid giant axon (non-myelinated) and for the sciatic nerve fibers (myelinated) of the frog, the results have been substantially the same. The protoplasm replacement experiments of Baker, et al., which are the keystone of the discussion, are the most recent, having been published in 1962. The relatively well time resolved experiments on heat evolution and absorption are also very recent. It is only these recent experiments which have permitted the assembly of experimental work on axon behavior into a remarkably self consistent pattern.

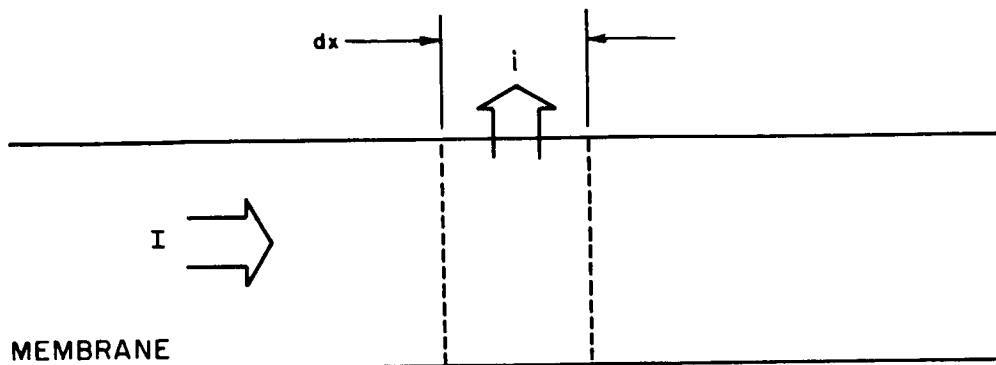
If a useful, self consistent description of nerve axon activity exists, it is reasonable to expect that the theory can be cast into the form of at least a phenomenological mathematical theory. This has been done with considerable success by Hodgkin and Huxley with special reference to the squid giant axon, (Ref. 6-10).

As is frequently the case in physical problems, the mathematical description of the nerve action potential begins with the derivation of a differential equation. The derivation can be made with the assistance of Figure 6-10.

Consider the membrane length dx . Let V_0 be the potential of the interior in the resting state and V be the difference between the potential at position x and the resting potential. There is an outward radial current flow per unit length of i , and the interior of the nerve has a resistance per unit length of r . If I is the longitudinal current, then Ohm's law gives

$$I = \frac{1}{r} \frac{\partial V}{\partial x} \quad (6-1)$$

FIG. 6-10
DERIVATION OF CABLE EQUATION FOR A NERVE



From Kirchoff's law, there is an outward flow of current,

$$i = \frac{\partial I}{\partial x} \quad (6-2)$$

Combining these equations yields

$$i = \frac{1}{r} \frac{\partial^2 V}{\partial x^2} \quad (6-3)$$

This is the equation used by Hodgkin and Huxley (Ref. 6-10) although their derivation is slightly different. Noting that

$$\frac{\partial I}{\partial x} = \frac{\frac{\partial I}{\partial t}}{\frac{\partial x}{\partial t}} \quad (6-4)$$

and that $\partial x / \partial t = v$, the velocity of propagation of a disturbance, one obtains

$$i = \frac{1}{r v} \frac{\partial^2 V}{\partial t^2} \quad (6-5)$$

These derivations are given by Tasaki (Ref. 6-11).

Now i is the current through the membrane and is independently determined by the potential difference across the membrane, the various ionic conductances and the capacitance. To develop their theory, Hodgkin and Huxley assume that the membrane has the characteristics shown by the model illustrated in Figure 6-11. There is a capacitance. In parallel with the capacitance are three series circuits. The first circuit is a variable sodium resistance in series with a "sodium pump". The second circuit is, similarly, a potassium pump in a series with a variable resistance to potassium ion flow, and the third is a similar arrangement for all other ion flows. The only justification for such an arrangement is agreement with experiment. From a logical standpoint, experimental agreement does not demonstrate that the model corresponds to reality, but only that it is one of a set of models which would correspond with reality. The function of a fundamental theory would be to derive this, or a similarly effective model, from considerations of molecular structure and chemical reactions. If,

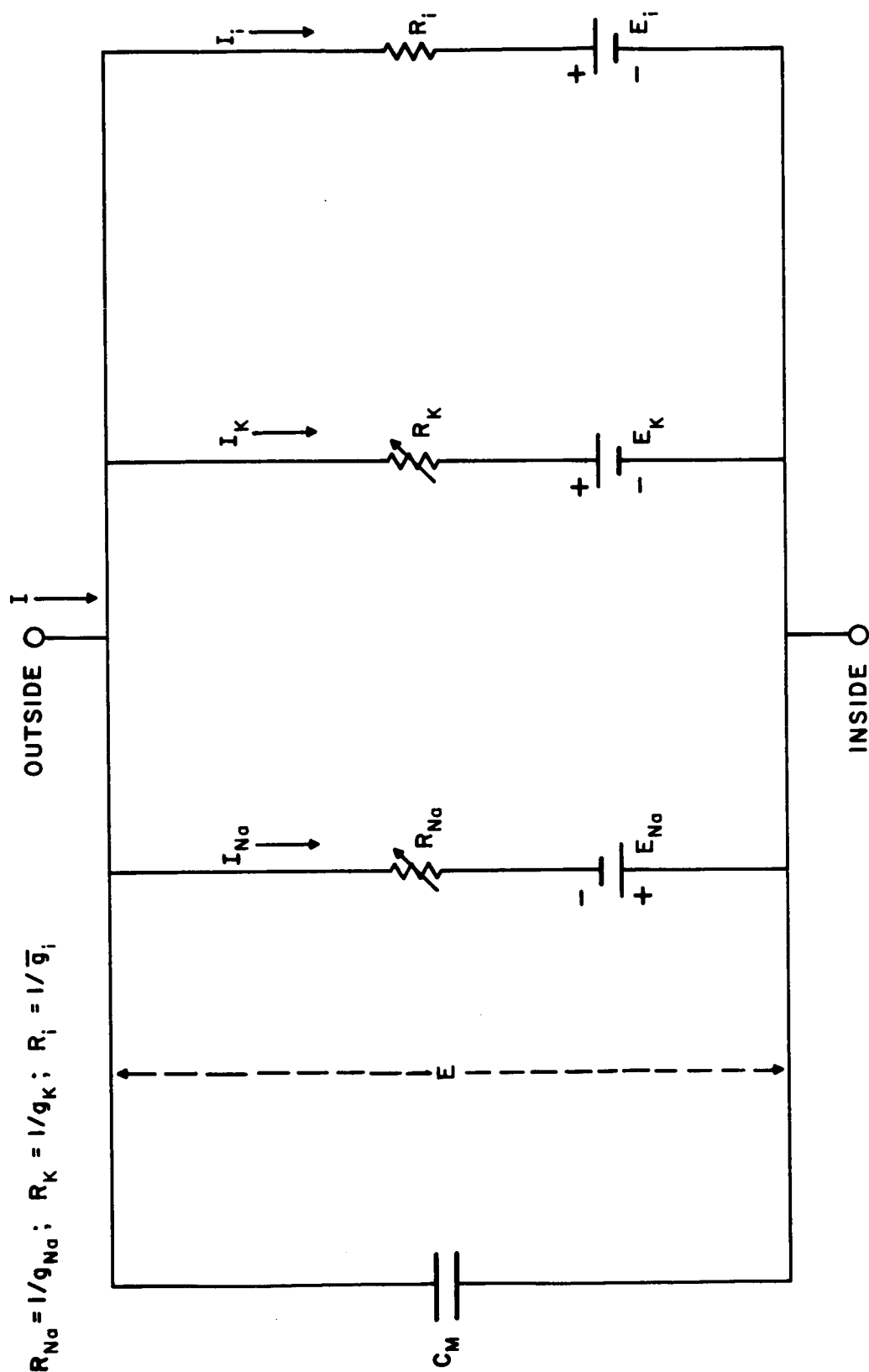


FIG. 6-11

ELECTRICAL CIRCUIT REPRESENTING MEMBRANE

038

however, the model actually explains a range of data greater than that originally contemplated in its construction, one is legitimately inclined to ascribe greater merit to the model.

The effective voltages of the pumps are E_{Na} , E_K and E_i for the sodium, potassium and other ions respectively. The ionic resistances are the reciprocal of conductances which are designated g_{Na} , g_K and g_i respectively. Experiments indicate that E_{Na} , E_K , E_i , C_m and g_i are independent of time and membrane potential difference. Whether or not an ion flow through the membrane occurs, and the flow direction, depend on the relationship between the membrane potential difference, E , and the respective ion pump potentials.

One can write formally,

$$I_{Na} = g_{Na} (E - E_{Na}) \quad (6-6)$$

$$I_K = g_K (E - E_K) \quad (6-7)$$

$$I_i = g_i (E - E_i) \quad (6-8)$$

where I_{Na} , I_K and I_i are the ionic currents due to sodium, potassium, and other ions respectively.

If one introduces a reference potential E_r and writes:

$$V = E - E_r \quad (6-9)$$

$$V_{Na} = E_{Na} - E_r \quad (6-10)$$

$$V_K = E_K - E_r \quad (6-11)$$

$$V_i = E_i - E_r \quad (6-12)$$

then Equations 6-6 to 6-8 become

$$I_{Na} = g_{Na} (V - V_{Na}) \quad (6-13)$$

$$I_K = g_K (V - V_K) \quad (6-14)$$

$$I_i = g_i (V - V_i) \quad (6-15)$$

It is possible to write formally:

$$g_K = \bar{g}_K n^4 \quad (6-16)$$

$$g_{Na} = m^3 h \bar{g}_{Na} \quad (6-17)$$

where n , m and h are variables between 0 and 1, and \bar{g}_K and \bar{g}_{Na} are constants.

If then one writes

$$\frac{dn}{dt} = a_n (1 - n) - \beta n^n \quad (6-18)$$

$$\frac{dm}{dt} = a_m (1 - m) - \beta m^m \quad (6-19)$$

$$\frac{dh}{dt} = a_h (1 - h) - \beta h^h \quad (6-20)$$

one has a formal set of equations to describe the membrane current.

The membrane current is then given by:

$$i = C_m \left(\frac{dv}{dt} \right) + \bar{g}_K n^4 (V - V_K) + \bar{g}_{Na} m^3 h (V - V_{Na}) + \bar{g}_i (V - V_i) \quad (6-21)$$

where the first term is the capacitance term.

This formalism was suggested by physical models but does not depend on the models.

The point of the discussion is that Hodgkin and Huxley used the results of voltage clamp experiments to obtain relatively simple analytic expressions for the α 's and β 's. Combination of Equations 6-21 and 6-3 yields a differential equation for the nerve impulse.

The formalism correctly gives the total ion current across the membrane in voltage clamp experiments.

A membrane action potential is defined as one in which the membrane potential is uniform, at each instant, over the whole length of the fiber considered and the net membrane current must, therefore, always be zero except during stimulus. If the stimulus is a short shock at $t = 0$, the form of the action potential should be given by solving Equation 6-21 with $i = 0$ and the initial conditions that $V = V_0$, and m , n and h have their resting steady state values. The formalism correctly describes this phenomenon.

Another test of the formalism is to solve Equations 6-21 and 6-3 together numerically, requiring that the resting potential be restored after passage of the action potential. When this is done, the velocity of the action potential is in substantial agreement with experiment. The results of these and other tests are summarized by Hodgkin and Huxley.

" The formalism was used to predict the quantitative behavior of a model nerve under a variety of conditions which corresponded to those in actual experiments. Good agreement was obtained in the following cases:

- (a) The form, amplitude and threshold of action potential under zero membrane current at two temperatures.
- (b) The form, amplitude and velocity of a propagated action potential.
- (c) The form and amplitude of impedance changes associated with an action potential.
- (d) The total inward movement of sodium ions and the total outward movement of potassium ions associated with an impulse.
- (e) The threshold and response during the refractory period.

- (f) The existence and form of subthreshold responses.
- (g) The existence and form of an anode break response.
- (h) The property of the subthreshold oscillations seen in cephalopod axons."

"The theory also predicts that a direct current will not excite an axon if it rises sufficiently slowly."

"Of the minor defects, the only one for which there is no fairly simple explanation is that the calculated exchange of potassium ions is higher than that found in Sepia (a squid closely related to Loligo) axons."

"It is concluded that the responses of an isolated giant axon of Loligo to electrical stimulation are due to reversible alterations in sodium and potassium permeability arising from changes in membrane potential."

This mathematical formalism and the associated experimental information discussed above defines the present degree of understanding of the nerve conduction process. The value of such a formalism is, of course, that it suggests further useful experiments. Hodgkin and Huxley in their paper mentioned one such experiment, namely, the influence of temperature on conduction velocity. They found that the formalism could be adapted to different temperatures simply by assuming that all the α 's and β 's had a Q_{10} of 3.* This immediately suggests another investigation. It is known that local temperature changes can trigger action potential propagation, (Ref. 6 -12). An interesting investigation would be to introduce time dependent changes in the α 's and β 's, corresponding to temperature changes, to determine if excitation is to be expected.

Sodium ionic movements are important in the phenomenological theory. This has led, in fact, to a misnomer: the theory is sometimes called the sodium theory. This is a misnomer because undue emphasis is placed on the importance of sodium: potassium ions are at least as important. Experiments have actually been performed in which sodium is replaced by different ionic species in the solution in which the nerve is immersed. In many cases, it has still been possible to conduct action potentials. Koketsu, (Ref. 6 -13) has argued that this demonstrates that the function of sodium is to affect some property of the membrane and

* Q_{10} of a current is the factor by which the current is increased for a temperature change of 10°C .

that in particular nerves some other ionic species are able to fulfill this role. In the mathematical model, however, the sodium ions and their migrations through the membrane are fundamentally important. A possible way to resolve this apparent contradiction would be to drop the sodium ion terms from the mathematical formalism and perform the calculations to determine whether or not there is an action potential and its form.

The role of calcium has also been placed in perspective using the formalism (Ref. 6-14). It has been observed that the effect of changes in calcium concentration is to change the values of the α 's and β 's in the formalism. These changes can be measured in voltage clamp experiments, and the formalism has then been used to calculate the dependence of nerve behavior on calcium concentration.

The objective of the discussion of this section has been to demonstrate that the action potential is understood in the sense that there exists a vast amount of data which can be integrated into a self consistent structure which can be expressed in a mathematical formalism and that the formalism can be a valuable tool in further research. Problems remain to be solved. The measure of progress that has been made is that it is now known what the problems are.

Electrical Excitation Experiments

Electrical excitation of nervous tissue is a necessary part of normal pulse transmission along axons. Electrical excitation is also experimentally convenient. The purpose of this section is to review classical experiments of this type and to discuss an in-house experiment performed at Allied Research Associates.

The effectiveness of excitation electrodes in exciting nerve impulses depends on the distance between electrodes. The classical experiment is that of Rushton, (Ref. 6-15). The experimental group setup is shown in Figure 6-12. Excitation of frog sciatic nerve was detected by twitch of the attached gastrocnemous muscle. The measurement performed was to determine the minimum potential difference between electrodes which causes muscle twitch as a function of interelectrode distance. Rushton used a pulse duration of 0.2 msec. In Figure 6-13, the results are plotted. This data point curve of threshold potential (which is presumably the meaning of Rushton's word "strength") versus interelectrode distance is plotted using certain normalizing techniques described by Rushton. The curve shape is typical. It is characterized by a general decrease of threshold potential differ-

FIG.6-12

RUSHTON'S APPARATUS FOR STRENGTH-LENGTH DETERMINATION

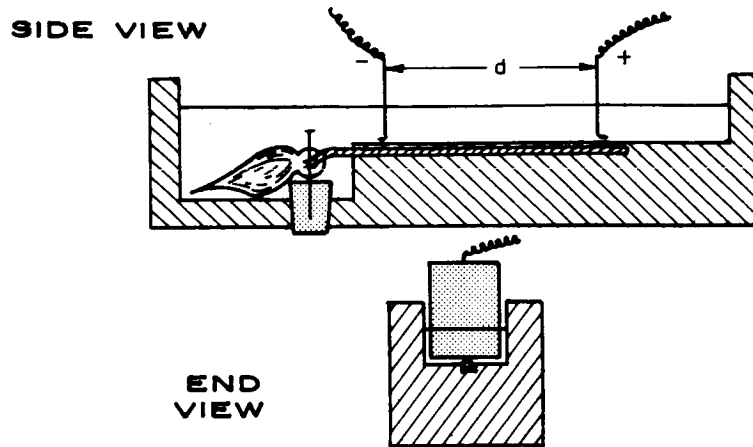
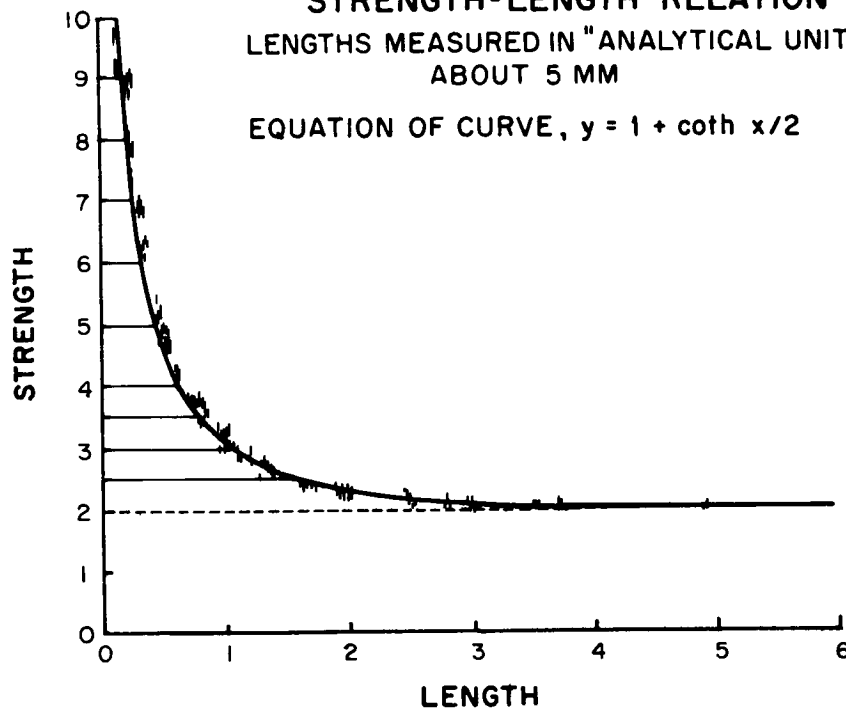


FIG. 6-13

STRENGTH-LENGTH RELATION

LENGTHS MEASURED IN "ANALYTICAL UNITS" OF
ABOUT 5 MM

EQUATION OF CURVE, $y = 1 + \coth x/2$



ence with increasing interelectrode distance. The data points seem to approach the zero ordinate asymptotically, although the data are necessarily not good enough to establish this with certainty. The curve also asymptotically approaches a constant nonzero value of threshold potential for large interelectrode distances. The solid curve shown is the result of a theoretical calculation which predicts a curve having the form:

$$y = 1 + \coth \left(\frac{x}{2} \right) \quad (6-22)$$

The threshold potential also depends on the duration of the excitation. The typical form of this dependence is shown in Figure 6-14, (Ref. 6-5). There is a minimum threshold potential for which excitation is impossible for any duration of excitation pulse. The plateau for which the threshold becomes nearly independent of duration extends from roughly 0.5 to 1.0 msec. to larger values for the frog sciatic nerve. In experiments at Allied Research Associates excitation with reasonable threshold potentials has been achieved for pulse durations as short as 0.05 msec. It is probable that the curve approaches the ordinate asymptotically although this should be investigated more carefully. The general shape of this curve is compatible with the results of Tasaki's experiments on depolarization at a single Ranvier node (Ref. 6-11).

The value of the threshold potential on the curve plateau is called the rheobase. It is convenient also to have a numerical designation which gives information on the shape of the curve. Many physiologists use the term "chronaxie" which is defined as the pulse duration corresponding to a threshold potential double the rheobase. This is not a very good shape designation but it is satisfactory for the present discussion. It was early observed that chronaxie increases with increasing interelectrode distance. Rushton attributes this to diffusion effects. "Ions are concentrated against the (nerve) sheath owing to the applied electric field; the concentration is dissipated owing to the concentration gradient and the potential gradient developed by the redistribution of ions", (Ref. 6-16). To the two dissipative effects cited by Rushton a third may be added: thermal diffusion. These dissipative effects are more effective for longer excitation pulses and this

FIG. 6-14
STRENGTH-DURATION CURVE

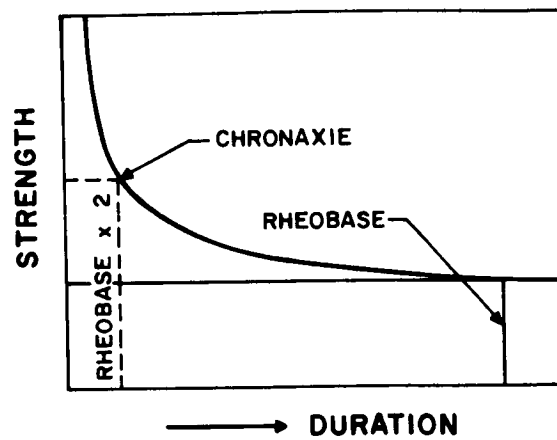
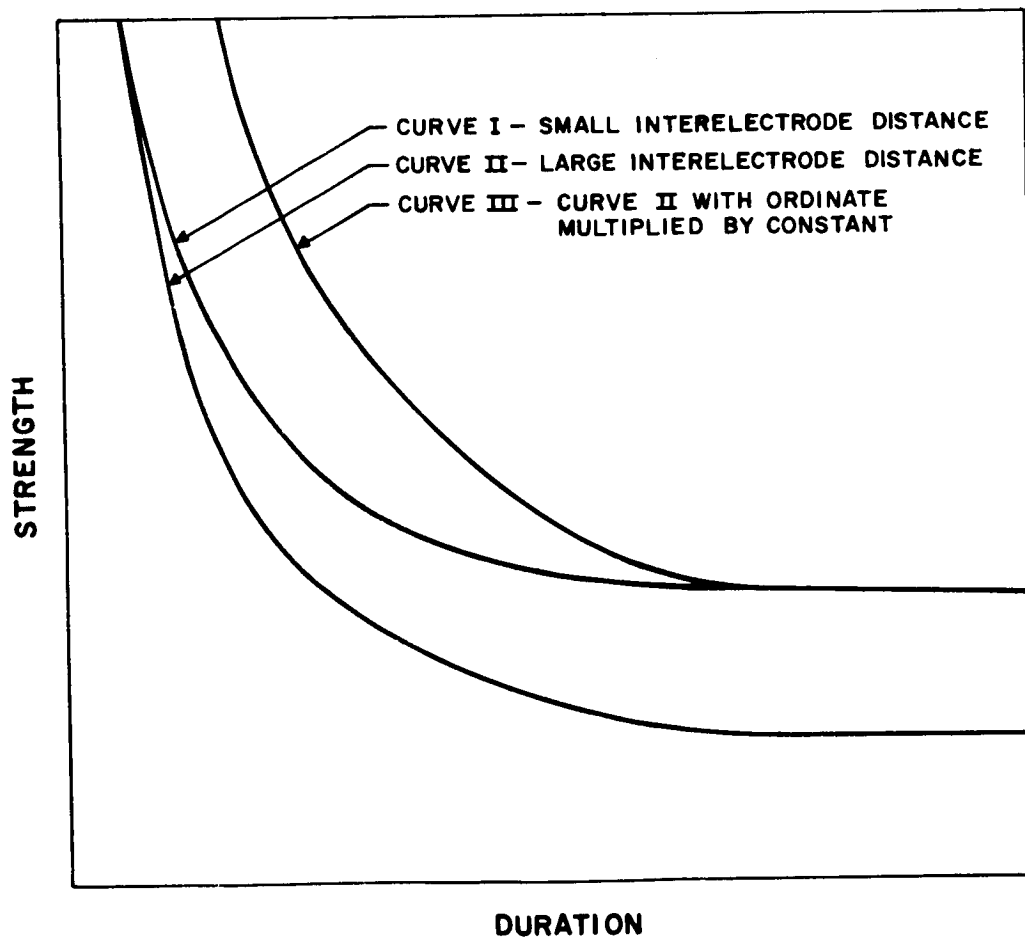


FIG. 6-15
DEPENDENCE OF CHRONAXIE ON INTERELECTRODE DISTANCE



time factor is the essence of Rushton's explanation. Rushton demonstrates convincingly that the dissipative effects are also more important for smaller interelectrode distances.

In Figure 6-15, let curve I represent the curve of threshold voltage versus pulse duration for a small interelectrode distance which is still great enough to be in the plateau region of Figure 6-13. Consider now an increase in interelectrode distance. Since diffusion effects are relatively unimportant for short pulses, the excitation curve for this case, curve II of Figure 6-15, will coincide with curve I for short pulse durations. For longer pulses, dissipation effects are important but less important than for curve I. The rheobase for curve II is less than for curve I.* Comparison of the two curves shows that curve II, for greater interelectrode distance, has a greater chronaxie than curve I. This can be seen by multiplying the ordinate of curve II by a constant to give curve III.

For electrode intervals shorter than those for the plateau, the excitation voltage increases with decreasing electrode interval. One might, therefore, expect that chronaxie dependence would reverse because the differences in electric field strength would dominate over the difference in the importance of dissipative effects. But these short interelectrode distances are precisely those for which the dissipative effects are most important and, evidently, predominate.

Dissipation effects can be important if they are actually responsible for chronaxie dependence on interelectrode distance because excitation threshold can be changed by factors of two, or even more, for pulse widths approximately equal to chronaxie.

The voltage clamp experiments described in Section 6.2.1, under "General Experiments and Theory", and the electrical excitation experiments of this section appear to form a consistent pattern of results.

* Strictly speaking, the interelectrode distances do not correspond to the plateau of Figure 6-15 as assumed since the threshold for greater interelectrode distance is less. Extending this argument leads to the conclusion that the plateau exists for interelectrode distances for which dissipation effects are either negligible or independent of interelectrode distance.

The voltage clamp experiments do not, however, deal with actual impulse excitation while the excitation experiments do not attempt to control the relationships among the several electric field components. It is reasonable, therefore, to design a set of excitation experiments in which a comparison is made among different sets of electrodes in which each set establishes a different electric field configuration. If the concepts assumed or implied in the above discussions of experiments are correct, these new experiments should yield no surprising results. Allied Research Associates has performed experiments, the results of which are sufficiently interesting to be discussed here. It is to be emphasized that these experiments are preliminary in nature and actually raise questions to be answered in further investigations.

For a preliminary investigation it was considered desirable to perform a standard physiology laboratory course experiment (Ref. 6-17) with small, but significant variations. The experimental setup is shown in Figure 6-16. A block of paraffin was hollowed out to form a box. Seven platinum electrodes were inserted through the side of the box. A square wave pulse whose duration and voltage could be adjusted over wide ranges was applied to the excitation electrodes. The pulse could be observed on one beam of a dual beam oscilloscope. The detection electrodes were connected to the terminals for the other oscilloscope beam, permitting observation of the action potential.

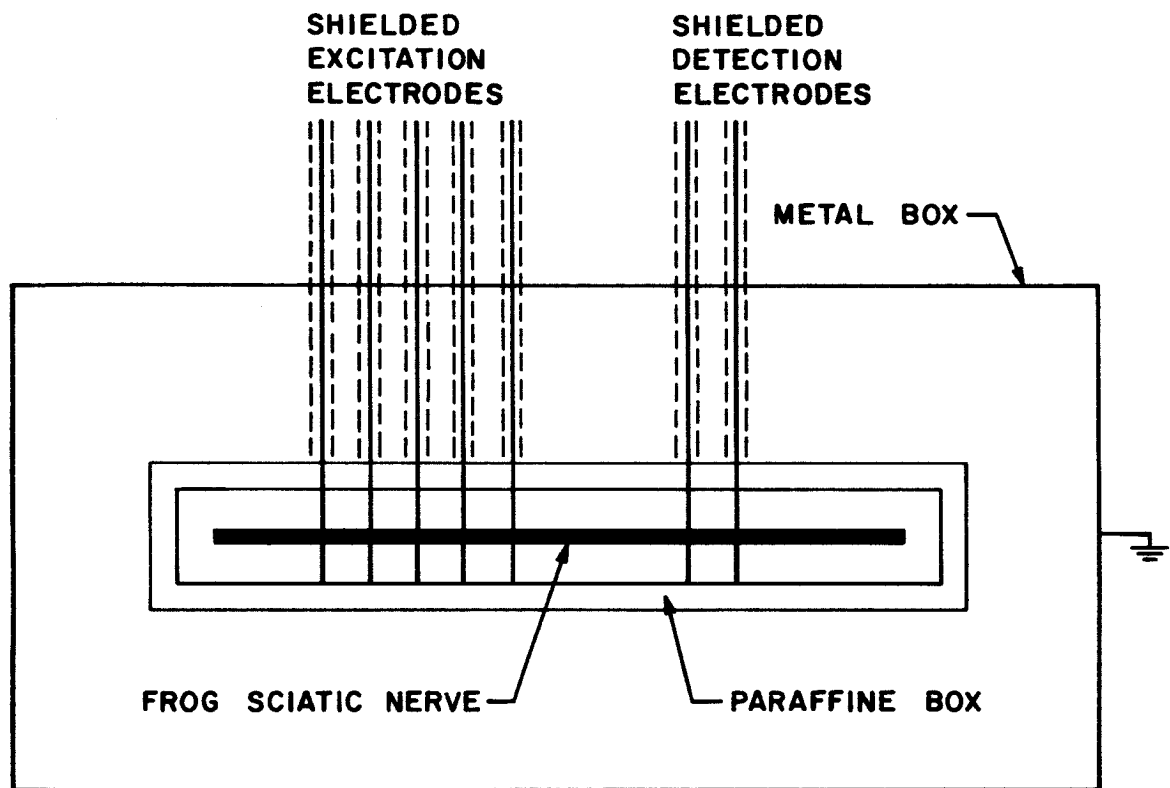
For a particular set of measurements, the sciatic nerve of a frog was laid across the electrodes after equilibration in Ringer's solution. Different excitation electric field configurations were obtained by connecting electrodes in different ways.

The nerve changes during a sequence of measurements. This requires careful planning to obtain the comparison sought in a particular set of measurements. If the shape of the excitation threshold versus pulse duration curve, that is, chronaxie, is of interest, the electrode configuration is kept fixed. The measurements are then performed with increasing pulse widths and then with decreasing pulse widths. Comparison between first and last measurements show how much the nerve has changed. If direct comparison between different field configurations is desired the pulse width

FIG. 6-16

APPARATUS FOR ELECTRODE CONFIGURATION EXPERIMENTS

(TOP VIEW)



METAL BOX CONNECTED TO GROUND
ALL ELECTRODE LEAD SHIELDS CONNECTED TO GROUND
ALL POSITIVE EXCITATION ELECTRODES CONNECTED TO GROUND

is kept fixed and the electrode connections are changed.

Changes in the nerve also shift emphasis from absolute measurements to relative measurements. This limitation can to some extent be overcome by more sophisticated methods contemplated for future investigations. Measurements were performed at room temperature; this and other variables would be more carefully controlled in future work.

Threshold potential difference is not a simple concept, especially for nerves which contain many fibers. A certain degree of judgment is necessarily involved. In the measurements discussed here two approaches were followed. In the first approach, threshold was considered to be the excitation voltage for which the first small perturbations in the detection electrode oscilloscope trace could be observed. This involved the excitation of very few fibers. In the second approach, a fully developed nerve action potential was used and threshold was defined as the voltage which produced a specified vertical displacement from the minimum between the α and β fiber maxima and either the α or β fiber maxima. This involved excitation of many fibers. It was found that if a particular threshold criterion was used consistently in a set of measurements, the relationships discussed below were independent of the particular threshold criterion selected. Unless otherwise noted, the comparisons described below were observed at least twice with no exceptions.

Four electrode configurations were of most interest although several others were investigated with results consistent with those of the four configurations discussed here. The configurations are shown in Figure 6-17. + represents a positive electrode, - represents a negative electrode and x represents an unconnected electrode. The electric field components normal to the nerve membrane and longitudinal to the nerve are sketched. Field fringing effects are estimated.

The threshold excitation voltages are designated by V_I , V_{II} , V_{III} and V_{IV} for the first, second, third and fourth configurations respectively. Chronaxies are similarly designated by C_I , C_{II} , C_{III} and C_{IV} .

FIG. 6-17

EXCITATION ELECTRODE AND ELECTRIC FIELD CONFIGURATIONS

+ Positive Electrodes

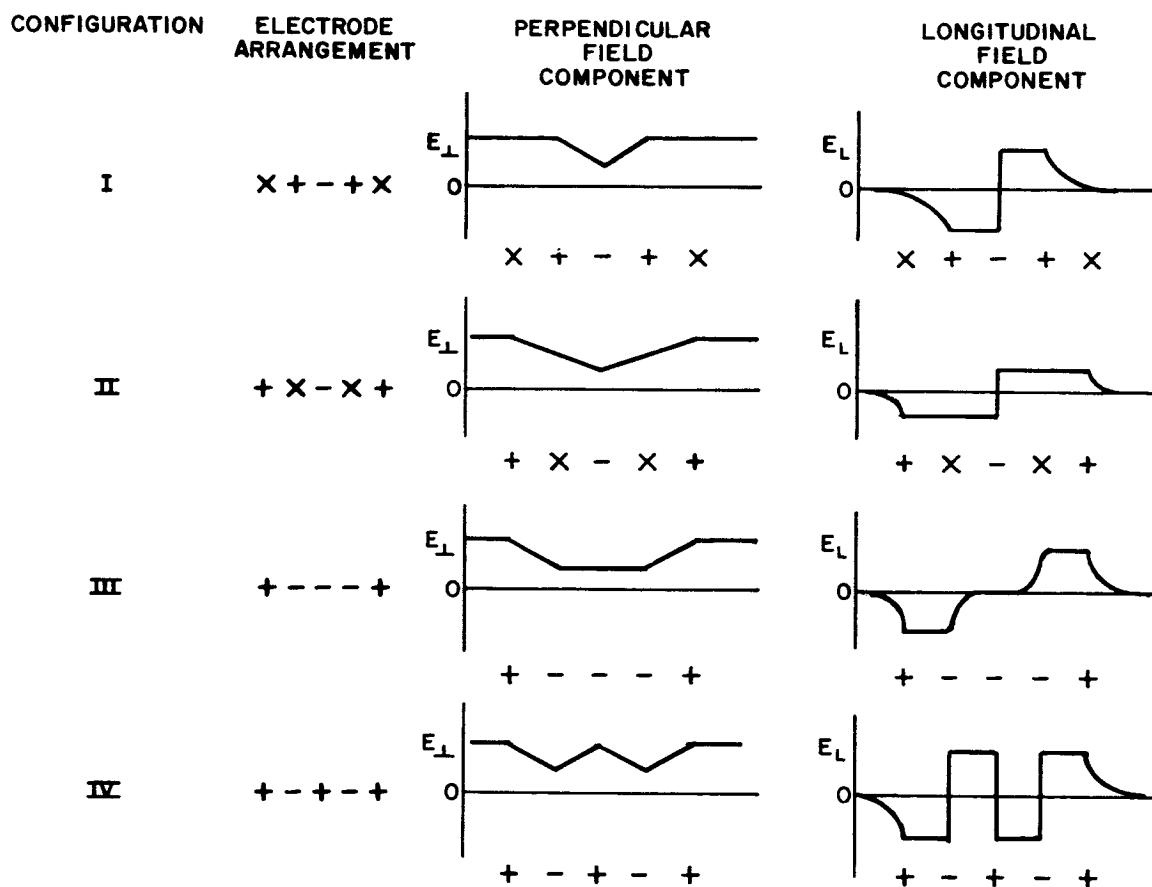
- Negative Electrodes

x Unattached Electrodes

E_{\perp} Electric Field Perpendicular to Membrane

E_L Electric Field Parallel to Longitudinal Nerve Axis

Note: Electric field sketches are crude estimates of general spatial dependence in steady state.



With this background, the experimental results can be briefly stated as follows:

1. $V_I \geq V_{II}$ (exceptions noted with one nerve)
2. $C_I < C_{II}$
3. $V_{III} < V_{II}$
4. $C_{III} < C_{II}$
5. $V_{IV} < V_{III}$ (nerve in air)
6. $V_{IV} > V_{III}$ (nerve in Ringer's solution, experiment performed with only one nerve)

There is nothing unusual in the first three items listed above. The electrodes are far enough apart, 12.5 mm, that the equality of V_I and V_{II} implies that for the particular nerve the plateau region of the classical curve of Figure 6-13 is involved. The dependence of chronaxie on interelectrode distance indicated by item 2 has already been noted. In configuration III, a larger electrode region is made negative than in configuration II. It is reasonable, therefore, to expect that potential difference across the nerve will be reduced more rapidly in configuration III than configuration II, hence item 3.

Items 4, 5, 6 do require explanation. In the discussion of Rushton's theory it was noted that large chronaxie corresponds to relatively small dissipation effects. The importance of dissipation effects can be considered proportional to a quantity having units of length, specifically, the average dissipation distance for ions during the time interval of interest. If $C_{III} < C_{II}$, then dissipation effects are more important in configuration III than configuration II, that is, that for configuration III, dissipation length is relatively larger than some significant linear dimension. It is clear that the nerve length under negative electrodes is not the linear dimension

because that is greater in configuration III. The best candidate is actually the nerve length for which there is an appreciable longitudinal field as well as a reduction in the perpendicular field (see Figure 6-17).

This focuses attention on item 5 in the list of observations. This item is at once the most unusual result as well as the result which was most carefully investigated in the measurements.* One would expect that configuration III would be most effective in reducing rapidly the potential difference across some portion of the nerve membrane but that configuration IV would be certainly no more effective than configuration I. Since configuration I is, experimentally, less effective than configuration II, one would expect that V_{IV} would probably be the greatest. V_{IV} is actually the smallest for the dry nerve.

Configuration IV yields longitudinal electric fields over a greater length of nerve than configuration I and III. The nerve length involved is the same as for configuration II but in configuration II the longitudinal field is weaker.

These results suggest that regions in which there is both a longitudinal field and a reduction in perpendicular field are regions of unusual interest in nerve excitation.

If this conclusion is correct, one would expect that shorting out the longitudinal field would appreciably raise the threshold in configuration IV. This can be done by immersing the nerve and the excitation electrodes in solution. Item 6 in the list of observations shows that the immersion actually does radically increase the threshold.

Two objections can be raised to the above discussion. The first is that in configuration IV, electrotonous effects are summed to produce the reduction in threshold. This could possibly account for $V_{IV} < V_I$ and $V_{IV} < V_{II}$. (It does not account for $V_{IV} < V_{III}$.)

The second objection is that the low values of V_{IV} are only observed in air but that nerves in vivo are immersed in solution. The

* Actually, configuration IV was investigated at first only for the sake of completeness: there appeared to be only one reasonable expectation, which was not supported by experiment.

essence of the matter, however, is that nerve activity in vivo involves very localized currents and this is true also of isolated nerves in air. For the isolated nerve immersed in solution, the current flows are, geometrically speaking, on a scale large compared with the diameter of the nerve. It would seem, therefore, that the measurements in air are actually more characteristic of in vivo conditions than the measurements in solution. This, of course, raises experimental difficulties because nerves in air degenerate more rapidly.

The preliminary experiments on the effect of electrode configurations discussed here do not, and were not intended to solve the problem of the importance of longitudinal electric fields. They do demonstrate that the subject is worthy of further investigation and that nerve activity cannot be understood until the issues raised here are resolved by further experiments.

It is premature to speculate whether the longitudinal field effects which seem to exist are related to the membrane or to local ionic flows outside of the membrane. If it should happen that the former is the case, this would have great significance for the membrane theory discussed in Sections 6.1, under "General Experiments and Theory", and 6.2.2.

6.2.2 The Axon Membrane

The discussions of Section 6.2.1 demonstrate that the key to the process of axon conduction is the behavior of the membrane. When the potential difference across the membrane is reduced slightly, the permeability increases radically and rapid depolarization results. Some process then reduces the permeability and restores the original polarization. This non-linear behavior is not, however, restricted to axon membranes and applies to other membranes. It is the subject matter of this section.

Von Neumann (Ref. 6.18) has made an important observation. "Since all these effects (relating to axon conduction) occur on a molecular scale - the thickness of the cell membrane is of the order of 10^{-6} cm which is a molecular dimension for the large organic molecules that are involved here - the above distinctions between electrical, chemical and mechanical effects are not so definite as they might at first appear. Indeed, on the molecular scale there are no sharp distinctions between all these kinds of changes: every chemical change is induced by a change in intramolecular forces which determine changed relative positions of the molecules, i. e., it is mechanically induced. Furthermore, every such intramolecular mechanical change alters the electrical properties of the molecule involved, and therefore induces changed electrical properties and changed relative electrical potential levels. To sum up: on the usual (macroscopic) scale, electrical, chemical and mechanical processes represent alternatives between which sharp distinctions can be maintained. However, on the near-molecule level of the nerve membrane, all these aspects tend to merge. It is, therefore, not surprising that the nerve impulse turns out to be a phenomenon which can be viewed under any one of them".

A further word of caution should be added: for the reasons given by Von Neumann, it may be possible to view the nerve impulse from each of these aspects but it does not necessarily follow that any one of these aspects, or any combination of them, will be adequate. This view is expressed in Schmitt's remark, "Until biophysics formulates concepts of its own, distinct from its parent (sciences), it is unlikely that any major advances will come from the field that could not equally well come from simple cooperation of biologists and physicists". (Ref. 6.19)

The difficulties of applying macroscopic physical concepts to biological phenomena are will illustrated by the nerve membrane.

It has only been quite recently that the existence of a membrane with a definite molecular structure has been unambiguously demonstrated by use of the electron microscope. Robertson's electron microscope work was first published in 1957 (Ref. 6-50). Danielli and Davson (Ref. 6-51) actually proposed much earlier a hypothetical membrane structure on the basis of permeability studies and experimental observation, but the electron microscope work was required to verify the general features of their hypothesis.

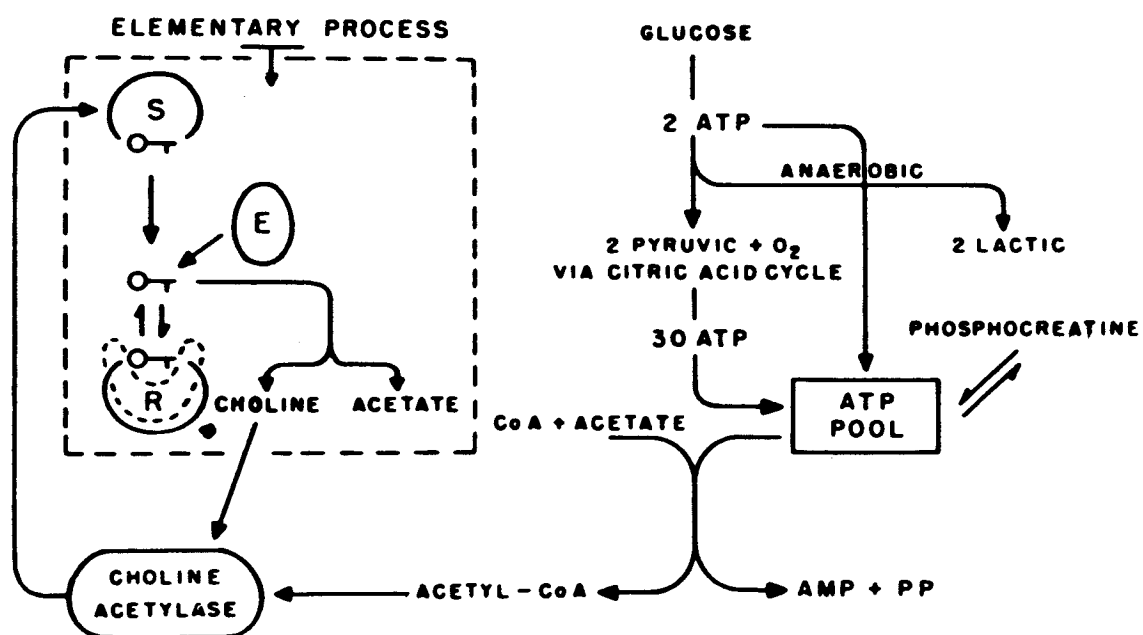
The importance of understanding nerve membrane behavior has been emphasized in the previous section. The question of what is meant by "understanding" arises. The behavior is known in the sense that resistance and capacitance have been measured as well as the changes which occur during passage of the action potential. Although the importance of longitudinal electric fields is not yet clear, the general conditions for triggering the changes in permeability are known. It appears that Nachmanson (Ref. 6.21) has identified another important factor in the triggering process. This is discussed below. The point to be made here is that one major thing remains, namely, to develop a mathematical theory which ties together these disparate pieces of knowledge. The purpose of this discussion is to attempt to identify the elements that might enter into such a theory.

Before proceeding further, it is necessary to discuss Nachmanson's theory of the role of acetylcholine in triggering the membrane permeability changes. His theory can be discussed with the help of Figure 6.18.

The role of acetylcholine in the elementary process may be pictured as follows: (1) In resting condition acetylcholine ($\text{O}-\text{T}$) is bound to a storage protein (S). The membrane is polarized. (2) Acetylcholine is released by current flow (possibly hydrogen ion movements) or any other excitatory agent. The free ester combines with the receptor protein (R). (3) The receptor changes its configuration (symbolized by dotted line). This process increases sodium conductance; it is the trigger action by which the potential primary source of EMF, the ionic concentration gradient, becomes effective and by which the action current is generated. (4) The ester-receptor complex is in dynamic equilibrium with free ester and receptor; the free ester is open to

FIG. 6-18

Sequence of energy transformations associated with conduction, and integration of the acetylcholine system into the metabolic pathways.



attack by acetylcholinesterase. (E). (5) The hydrolysis of the ester permits the receptor to return to its original condition. The sodium conductance decreases and the membrane returns to its original condition.

One must make a judgement of whether Nachmanson's theory is or is not substantially correct before speculating further on the operation of the membrane. The evidence advanced by Nachmanson, however much confusion there is over details, strongly supports his hypothesis.

It will be recalled from the last section that the action potential phenomena do not seem to require presence of the protoplasm. The acetylcholine cycle must take place within the membrane.

More than the acetylcholine cycle occurs in the membrane. This only is intended to explain the triggering mechanism which increases the sodium permeability to initiate the action potential. With this permeability increase sodium ions simply flow down their concentration gradient, i.e., depolarization occurs. After passage of the action potential the nerve is restored to its usual state. To do this, sodium and potassium ions must be moved against their potential gradients and polarization restored. This requires energy and some kind of mechanism. The phrase "sodium pump" is used to refer to the mechanism for sodium.

It is apparent, therefore, that the membrane is the scene of a complex combination of events which seems to require a variety of materials. Knowledge of the membrane structure is necessary before one can hope to understand the combination. It is to chemical, x-ray and electron microscope studies that one must turn for this knowledge.

The structure of the membrane as revealed by the electron microscope can be summarized in the following statement. "The membrane of the Schwann cell as well as that of the axon consists of a structure measuring 75 Å in total thickness and appears as a pair of strata each about 20 Å thick separated by a light central zone". (Ref. 6.3). Knowing that the resting potential during the passage of the action potential is approximately 100 mv, one can immediately perform an interesting calculation. The result is that the electric field across the membrane in the resting state is the order of 100,000 volts/cm and that the changes in electric field during passage of the action potential are the order of 150,000 volts/cm. If the potential difference

in the membrane is applied through a distance less than the membrane thickness, these fields would correspondingly be greater. Fields of these strengths are sufficient to distort molecular structures.

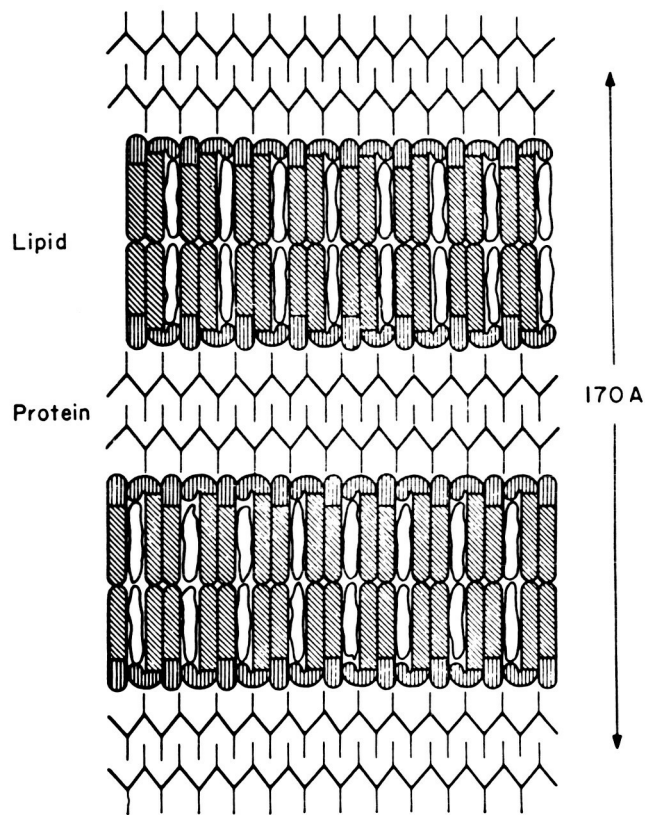
The macroscopic behavior of the membrane yields several valuable clues to the microscopic behavior. Nachmanson notes that during unimpaired passage of the action potential about 10^{-14} moles of acetylcholine are metabolized for each square centimeter of membrane, and that 4×10^{-12} moles of Na enter the nerve during the same event. He estimates that 500 to 1,000 Na ions cross the membrane for each molecule of acetylcholine metabolized. The above mole figures correspond to 0.6×10^{10} molecules of acetylcholine and 2.4×10^{12} molecules of Na ions. Assuming equal spacing, the acetylcholine molecules which are hydrolyzed are 770 \AA apart. Figure 6.19 gives a diagram of Finean's model of the repeating structure of myelin which should be similar to that of the membrane itself (Ref. 6.1). The repeating unit along the membrane has a dimension of about 13 \AA . Nachmanson estimates a single acetylcholine molecule to cover a membrane area of $30 - 50 \text{ \AA}^2$ or a diameter of $6 - 7 \text{ \AA}$. It is clear that triggering the acetylcholine cycle during passage of the action potential is an improbable event. If one denies the importance of acetylcholine and assumes equal spacing of the Na ions crossing the membrane, then the spacing is 69 \AA . Only about one repeating unit in twenty-five passes a sodium ion. This is still an improbable event.

It is interesting that as far back as 1932, A. V. Hill argued that molecular changes can only occupy a small fraction of the membrane surface (quoted in Ref. 6.21). The basis of his argument was that the heat evolved was such that the heat per molecule was several orders of magnitude less than the energy of a quantum of visible light. More recent heat evolution measurements discussed in Section 6.2.1 do not substantially change this argument.

In ionic conduction, the ionic velocity in a field of 1 v/cm is the mobility. Mobilities are the order of 10^{-3} cm/sec or 10^{-6} cm/msec . Because of the strong field across the membrane, it is apparent that if a Na ion can pass through at all it will do so in a time short compared with the duration of the action potential.

FIG. 6-19

Finean's conception of the molecular organization of the 170 Å repeat unit in nerve myelin.



It is evident that the current flowing through the membrane cannot be treated as a continuous flow. It is actually the summation of many sporadic, essentially improbable events. Such a process must be described in statistical terms, and it is desirable to seek information on the kind of statistics that may be involved.

One way to investigate this question is to determine the sodium ion current flow as a function of depolarization voltage. This can be done from the Hodgkin and Huxley Equation 6.13 (Section 6.2.1), using data in their paper (Ref. 6.1). There is one difficulty. If a constant depolarization voltage is applied across the membrane, the sodium ion conductance is given by

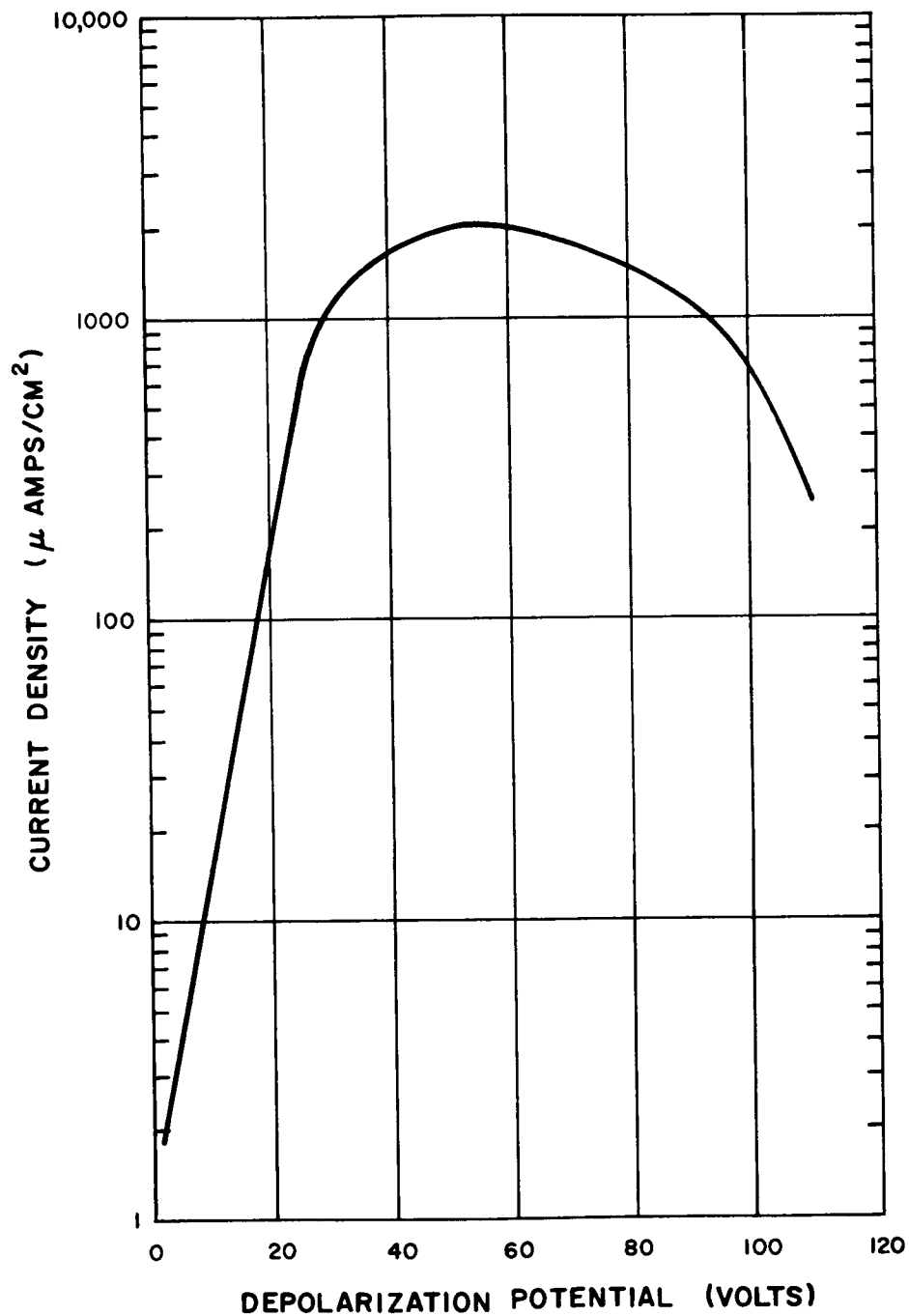
$$g_{Na} = g'_{Na} \left(1 - e^{-t/\tau_m}\right)^3 e^{-t/\tau_h} \quad (6-23)$$

where g'_{Na} , τ_m and τ_h are constants and t is time. It is clear that g_{Na} increases to a maximum and then approaches zero asymptotically. This behavior would be expected if the phenomenon were determined by two time dependent physical processes, one of which, acting alone, would tend to cause an increase in conductance and the other would tend to counteract the first. For the present discussion only the first process is of interest and the conductance to be used is the maximum conductance assuming the second process is ineffective. This maximum is obviously g'_{Na} and it is the value of the conductance which was used to calculate the currents in Figure 6.20.

Figure 6.20 is a semilog plot of sodium ion current density through the membrane as a function of depolarization potential. The first part of the curve is linear. The maximum occurs at approximately the resting potential and need not be regarded as surprising. It is of interest to calculate the number of ions passing through 1 cm^2 of membrane in one millisecond for the $1,000 \mu \text{ amp/cm}^2$ current. The charge passed is obviously 10^{-6} coulombs which corresponds to 6.28×10^{12} sodium ions per cm^2 . The spacing between charges is about 38 \AA . If the repeating unit in the membrane has a diameter of 12 \AA , about 10% of the units in the membrane pass an ion in one millisecond. If the acetylcholine cycle exists, the percentage of repeating units involved is radically smaller. The membrane current appears again to be the summation of essentially improbable events.

FIG. 6-20

SODIUM ION CURRENT DENSITY THROUGH
SQUID AXON MEMBRANE AS A
FUNCTION OF DEPOLARIZATION POTENTIAL



It is appropriate to apply Boltzmann's statistics to a simple potential well problem. Let it be assumed that the membrane is an ensemble of systems. Let the average energy of each system in the ensemble be kT and suppose each system is in a potential well E_o . Let N be the total number of systems and n be the number of systems whose energy is greater than E_o . Then Boltzmann statistics lead to the result that

$$n = Ne^{-E_o/kT} \quad (6.23)$$

It is not unreasonable to assume that the effect of reducing the membrane potential is to reduce E_o . In the acetylcholine cycle, for example, E_o would be the potential which holds the acetylcholine in the storage protein. Whatever the precise physical mechanism, it is not unreasonable to assume that the current through the membrane will be proportional to n .

If E_o is given by a function of the form $\beta (F - F_o)$ where F is the resting potential and F_o is the depolarizing potential,

$$n = N \left[e^{-\frac{\beta F}{kT}} \right] \left[e^{+\frac{\beta F_o}{kT}} \right] \quad (6.24)$$

Note that n increases for increasing depolarizing potential or decreasing membrane potential. The dependence is also exponential.

If $n \ll N$, then the height of the potential well, E_o , is considerably greater than kT . E_o is determined by the structure of the storage protein and of the acetylcholine. If E_o is large, considerable distortion of the structure would be required to change E_o , but it has already been noted that there are strong electric fields in the membrane.

Equation 6.23 also implies a temperature dependence. It is desirable to calculate what this should be. It is apparent from Equation 6.23,

$$\frac{n(T+\Delta T)}{n(T)} = \frac{e^{-\frac{E_o(1-\alpha\frac{\Delta T}{T})}{k(T+\Delta T)}}}{e^{-E_o/kT}} \quad (6.25)$$

if ΔT is an increment in temperature and the dependence of the potential barrier is a function of the form

$$E_o(T+\Delta T) = E_o(T) (1 - \alpha \frac{\Delta T}{T}) \quad (6.26)$$

with α being a constant.

If $\Delta T \ll T$, Equation (6.25) becomes

$$\frac{n(T+\Delta T)}{n(T)} = e^{+\frac{E_o}{kT} (\alpha \frac{\Delta T}{T} + \frac{\Delta T}{T})} \quad (6.27)$$

If the temperature dependence of the current, $i(T)$, is given by:

$$\frac{i(T+\Delta T)}{i(T)} = e^{\delta} \quad (6.28)$$

and if i is proportional to n , then,

$$\delta = (1+\alpha) \left(\frac{\Delta T}{T} \right) \left(\frac{E_o}{kT} \right) \quad (6.29)$$

$$\text{and, by definition, } \delta = (1+\alpha) \left(\frac{\Delta T}{T} \right) \left[\frac{\beta (F - F_o)}{kT} \right] \quad (6.29a)$$

This equation shows the relationship between the influence of depolarization potential on current through the membrane and the temperature dependence. To use this equation one must speculate about α . If, for example, α were 1, doubling the temperature would eliminate the potential barrier. This is not unreasonable. Fundamentally, of course, α is related to chemical bond energies.

Equation 6.29a implies that Q_{10} for current at a given temperature is a function of the depolarization potential.* Published data do not give satisfactory information on this. There is some evidence that the threshold membrane depolarization in the giant squid axon has $Q_{10} = 1.54$ (Ref. 6-22). Hodgkin and Huxley give $Q_{10} = 3.0$ for depolarization currents in general. Since the general phenomena would be more strongly influenced for strong, depolarization, and the threshold depolarization is relatively weak, these two values are consistent with Equation 6-29a. There is, of course, need for a consistent set of measurements.

Hodgkin and Huxley used a Q_{10} of 3.0 which corresponds to a δ of 1.10. Letting $\alpha = 1$, and remembering that $T \approx 300$, one obtains $E_o/kT = 16.5$.

* Q_{10} of a current is the factor by which the current is increased for a temperature change of 10°C .

This means that in Equation 6.23,

$$n/N = e^{-16.5} \quad (6.30)$$

$$n/N = 6.8 \times 10^{-8} \quad (6.31)$$

It was noted above that for a $1000 \mu\text{amp}/\text{cm}^2$ current, 6.28×10^{12} sodium ions are passed per cm^2/msec . This corresponds to activation of 1.26×10^{10} acetylcholine cycles assuming with Nachmanson that 500 sodium ions pass for each acetylcholine system activated. Also, assuming with Nachmanson that each acetylcholine system occupies 30 \AA^2 , the active area of 1 cm^2 of membrane during one millisecond is 3.78×10^{-5} . This is to be compared with 6.8×10^{-8} in Equation (6.31).

The agreement is actually good. The ratio n/N is very sensitive to Q_{10} , and Hodgkin and Huxley observed Q_{10} values ranging from 2.7 to 3.5. The value of α was only a guess. If $\alpha = 1.5$ and $Q_{10} = 2.7$, $n/N = 0.6 \times 10^{-5}$. A simple application of statistics thus yields a remarkable correlation of the experimental data.

These ideas should be pursued further. If the potential well concept is once established as acceptable, the problem becomes twofold. First, experimental results must be used to establish the parameters of the potential well. Second, the parameters must be correlated with the structure of the appropriate molecules. This second problem could be treated in two ways. The experimentally determined parameters can be used to draw conclusions about the chemical structures and knowledge of the structure can be used to calculate the parameters. The procedures involved in these problems are well established. There is ample precedent in both nuclear and solid state physics.

The preliminary problem worked out above does not involve the use of quantum mechanical concepts. This is not surprising since ions are sufficiently heavy particles that they would be expected to have small de Broglie wavelengths. Further investigations should, however, determine definitively whether the problem is classical or quantum mechanical.

It is believed that the above discussion demonstrates that a fundamental theory of the axon membrane lies in the near future.

6.2.3 The Logic of the Nervous System and Pattern Recognition

The organization of the nervous system is the subject of a vast amount of literature. Biological organisms detect and monitor their environment by means of sensors. These sensors are energized by the environment and supply information to the nervous system in the form of nerve impulses. The process of converting the patterns of nerve impulses into responsive modes of action is the logic of the nervous system. This logic is of fundamental interest to the bionicist.

The sensors as such are discussed elsewhere in this and previous (see Ref. 6-20) reports. The purpose of this section is first to isolate what appear to be general principles employed by biological organisms in the evolutionary design of sensor systems, and then to discuss the logic by means of which the nervous system converts these sensor signals into meaningful form. Only those references in the literature which are of immediate relevance are cited.

The first general principle to be discussed is Fechner's law. Fechner gave mathematical formulation to earlier work of Weber. He stated that the incremental response to a stimulus is related to the increment of the stimulus by the equation:

$$\frac{\Delta S}{S} = kR \quad (6-32)$$

where S is the stimulus

R is the response

k is a constant

Fechner integrated the Equation 6.32 to yield

$$R = a \log S + b \quad (6.33)$$

where a and b are constants.

Fechner's law was a purely empirical relationship which applies to most sensory receptors with varying degrees of precision. Granit (Ref. 6-23) remarks, "It would show lack of historical sense to look upon Fechner's famous integration as a mathematical treatment that could not

be improved upon". Reference to "historical sense" does suggest the possibility of a useful analogy. From a logical standpoint, Fechner's law is similar to one of Kepler's laws of planetary motion. It condenses a vast range of experimental data into a simple mathematical formula.

Perhaps the second Keplerian type law of receptor physiology is that due to Adrian who discovered that the "code" by which information is transmitted from receptors is based on nerve impulse frequency and that impulse frequency tends to be proportional to the log of the intensity of the stimulus.

The frequency coded impulses are in the nerve and the stimulus is applied to a receptor to which the nerve is attached. It is natural to investigate the connection between the two phenomena. In measurements, performed using frog muscle spindle, it was observed that impulse frequency was proportional to the logarithm of the load on the muscle, in accord with Fechner's law. It was also observed that the local depolarization occurred and that the frequency appears to depend in a linear way on depolarization. This would imply that the logarithmic dependence is involved in the mechanism that produces membrane depolarization. The investigation of this phenomenon has not been thorough enough to make a simple descriptive mathematical formulation possible. Such a description would be a third Keplerian type law.

The nerve attached to a receptor is not myelinated. The mathematical investigations of Hodgkin and Huxley discussed in Section 6.2.1 may therefore prove of fundamental importance in explaining receptor behavior. The logarithmic dependence of Equation 6-24 should be noted in this connection.

Pursuing the historical analogy somewhat further, psychophysics has not yet had its Newton.

The preceding discussion relates only to receptor units and their immediately attached nerve fibers. It is impractical in the present context to attempt a description of all known behaviors of the various types of receptors in biological organisms. Fechner's law and the related information is, however, remarkably general and is likely, therefore, to be of bionic significance.

Equally interesting to the bionicist is the organization of the signals leaving a set of parallel receptor units before the signals are transmitted to the

brain. Mammalian vision is a good example of this. The optic nerve fibers are the third layer of neurons in the passage from the retina to the brain. There are many more rods and cones in the retina than nerve fibers in the optic nerve. In peripheral vision there are complicated interconnections among the neurons in successive layers. In foveal vision there is a much closer approach to a one to one connection between the foveal cones and the optic nerve fibers. Granit, (Ref. 6.23) however, argues against a perfect one to one association, even in this case.

In general, an optic nerve fiber is connected with many rods and/or cones. The "domain" of a single optic nerve fiber may lie outside that of another, or may overlap another, or may be contained within another. The relationships are complicated and certainly not perfectly known. It seems clear, however, that the pattern of interconnections is important in organizing the information secured by the receptors into a form suitable for transmission to and interpretation by the brain. A rather remarkable fact is that skin receptors are similarly grouped into domains with complicated relationships. This appears to be a general organizational principle.

There are certain special cases of this organization principle which have engineering analogues. The distinction between foveal and peripheral vision is analogous to that between tracking and searching radar. If "domain" be generally interpreted as a range in color sensitivity, then certain color detection machines can be considered somewhat analogous to color perception in the eye.

Fiber optics* provides another illustration of the domain principle. The resolution of a bundle of optical fibers is limited by the diameter of the individual fiber. The resolution can be improved by projecting the image onto the ends of two bundles. This is so because a point on the image may not fall on the active part of the end of a fiber in one bundle but it may do so in the other bundle and thereby be observed.

*In an interdisciplinary report it is appropriate to caution the reader that fiber optics has nothing to do with the optic nerve. It does have to do with light pipes. An optical fiber is a thin light pipe.

It is appropriate now to turn to the operation of the nervous system as an integrated whole. It is not intended to survey the various schools of thought on this subject. Chauchard (Ref. 6.24), for example has done this very well. Von Neumann (Ref. 6.25) has written a classical discussion of the analogy between the brain and the computer. It is not appropriate to repeat or condense these discussions for the present purpose. There is, however, one aspect of the organization of the nervous system upon which it is useful to comment.

What does a bionicist expect to learn from examination of the operation of the nervous system? One can note that the human central nervous system has the order of 10^{10} neurons - many more elements than any computer - and that these neurons are packed more densely, by a factor of 10^8 to 10^9 , than the most compact computer elements. But the computer designer does not need to be told by the bionicist of the benefits of being able to use large numbers of components which have small volume and power requirements. If, in fact, the only substantial difference between the computer and the brain were the number of components, the analogy would not be at all interesting to the bionicist.

The deficiency and the strength of computers have been nicely stated by MacKay. (Ref. 6.26). "Every scientific subject has its 'honest hodmen' - men who excel in technique, and can turn out quantities of valuable results from a persevering and capable application of it. A conventional computer, however capacious and speedy, is the honest hodman per excellence. Like its human prototype, it earns respect and gratitude for indispensable services; but for all its dexterity and reliability, one thing is lacking. It is not 'bright'. It is indeed desparately less bright than the dullest hodman, for even he permits himself an occasional spontaneity when off duty. It is humanly impossible to be as dull as a digital computer".

The bionicist asks, therefore, whether or not there is some principle of organization used by the nervous system which can be incorporated into computers to render less true the last sentence of the above quotation.

One school of bionicists is attempting to answer this question by introducing statistical concepts and probability into computer operation. This is the approach of, for example, the designers of the perception (Ref. 6.27).

As more effort is expended useful results can be expected because the approach is valid, at least by analogy with the operation of the brain.

One does have a lingering suspicion, however, that if nature is able to construct 10^{10} computer components but unable to design components of good accuracy, then she is constrained to use randomness, probability and statistics in the design of biological computers. It is reasonable to enquire whether or not there is some more positive principle of computer design which nature uses and which is more readily adaptable in engineering designs. The remainder of this section will argue that there is, indeed, such a principle.

Von Neumann (Ref. 6.25) remarks, "The above description (of the neuron) contains some idealizations and simplifications, which will be discussed subsequently. Once these are taken into account, the digital character no longer stands out quite so clearly and unequivocally. Nevertheless, the (digital) traits emphasized in the above are the primarily conspicuous ones. It seems proper, therefore, to begin the discussion as done here, by stressing the digital character of the nervous system".

Von Neumann actually devotes considerable space to the question of the analogue properties of the brain. His remarks have been amplified elsewhere. His general opinion seems to have been, however, that there is no way of evaluating the importance of analogue operation and that speculation would not be very fruitful.

Recently the writer prepared an internal memorandum which sought to evaluate the importance of analogue procedures in mental operations. The approach in the memorandum is somewhat unconventional* and it reads as follows:

"Perhaps the most fascinating analogy in the field of bionics is that between the brain and the computer. It may also prove the most fruitful from the viewpoint of applications. If the bionics approach is to be followed, it is appropriate to analyze operation of the brain to serve as a guide for computer investigations.

*The memorandum was actually composed half in jest. Sometimes a jest can be quite serious.

"Two general lines of investigation of possible relevance to the present discussion have been pursued. The first is the science and philosophy of psychoanalysis. The second is, perhaps, of a more philosophical nature, especially studies by Hadamard and Poincare (Ref. 6.28)."

"The difficulty until recently has been that there has been little interaction between the two lines of inquiry. The reason for this has been the particular form taken by psychoanalytic theory. Certain modifications in this form proposed by Kubie (Ref. 6.29) permit integration of the two lines of inquiry. The purpose of this memorandum is to discuss this integration. An interesting result of the discussion will be to shed some light on certain questions raised by Wiener and Von Neumann (Ref. 6.30 and 6.25)."

"The issues revolve around Freud's terms conscious, subconscious and preconscious. Since Freud was primarily interested in mental illness his emphasis was on the conscious and the subconscious; the preconscious was more or less a gatekeeper between the two. This theory is effective in discussing mental illness but Kubie has pointed out it is much less useful in discussing higher mental operations. To Freud, the creative process resides in the unconscious. The mechanisms involved are unclear and, as is to be expected, Freud's discussions of the creative process are relatively sparse and not too satisfactory. He is most successful in discussing the motivation for higher mental activity but when the discussion of Hadamard is read, it is clear that something is missing. The contribution of Kubie is to show that in classical psychoanalytic theory the role of the preconscious is not sufficiently well defined. As a matter of fact, by the time Kubie finishes describing his picture of the preconscious it is clear that the Freudian picture of the preconscious is a very small part of Kubie's."

"Kubie begins by noting an interesting feature of Freudian theory, namely, that both the conscious and the subconscious operate using symbolism. Kubie claims that a large segment of the brain operates non-symbolically and to this part of the brain he gives the name 'preconscious'. He also asserts that higher mental processes, including creativity, consist of cooperation between the symbolically operating conscious and the non-symbolically operating preconscious. The symbolically operating subconscious may provide motivation but it is essentially inimical to effective thought."

"In view of the above discussion, how would one expect a creative act to occur? First of all, the problem would be defined consciously, perhaps under the influence of the unconscious. (In the case of the neurotic genius, one would expect that the influence of the unconscious would result in rather monotonous restatements of the same problem -- a common phenomenon.) Then the conscious works hard at the problem and fails to solve it. But the effort does activate the non-symbolic preconscious and provide it with data. Since the conscious is symbolic in its operation it is not aware that the problem is in process of solution until the solution is suddenly presented to the conscious. The solution must then be consciously verified."

"This is almost inevitably the way the brain works, with variations that are sufficiently minor that citing particular cases would be monotonous.

"What does all this have to do with computers? The analogy is straightforward. Digital computers are symbolic in their operation and analogue computers are non-symbolic. The brain operates like a combination of digital and analogue computers."

"Von Neumann (Ref. 6.25) and others have noted that the brain achieves unusually accurate results with imprecise elements. This, in a sense, is what analogue computers do. Wiener (Ref. 6.30) remarks that the brain seems to use scanning operations. In identifying a square, for example, the brain has 'square' set into it somehow and then changes scale until the preset 'square' in the brain matches the perceived square. This is an analogue operation."

"What has all this to do with bionics? This can be answered if one considers what the analogue elements probably are in the brain. It is probable that they are actually assemblies of digital elements, which nature can easily construct because of her skill in constructing small scale systems. But it is the analogue unit which results from the assembly of digital units which is functionally important, and it is in the realm of analogue units that human engineering has an advantage. If this is so, then man may be able to compete with nature by using a relatively small number of digital units to control a number of analogue units. The interesting area of study would, therefore, be in the 'interface' between the digital and analogue components."

Since preparation of the memorandum the writer has learned that there is a considerable school of thought which emphasizes the possible value of combining digital and analogue techniques in the same computer or device. This school is lead by MacKay (Ref. 6.26).*

The discussion of the memorandum goes further, however, than simply pointing out the combination of digital and analogue processes in mental operations. It explicitly identifies analogue operations with those aspects of mental activity which are most "mysterious", specifically, the creative, imaginative and insightful operations of the mind.

Rosenfeld (Ref. 6.32) in discussing the epistemology of Niels Bohr, the physicist, gives the following quotation written by Bohr when he was still a licentiate at the University of Copenhagen:

".....certainly I have seen before thoughts put on paper; but since I have come distinctly to perceive the contradiction implied in such an action, I feel completely incapable of forming a single written sentence. And although experience has shown innumerable times that it can be done, I torture myself to solve the unaccountable puzzle, how one can think, talk or write. You see, my friend, a movement presupposes a direction. The mind cannot proceed without moving along a certain line; but before following this line, it must already have thought it. Therefore, one has already thought every thought before one thinks it. Thus every thought, which seems the work of a minute, presupposes an eternity. This could almost drive me to madness. How could then any thought arise, since it must have existed before it is produced? When you write a sentence, you must have it in your head before you write it; but before you have it in your head, you must have thought it, otherwise how could you know that a sentence can be produced? And before you think it, you must have had an idea of it, otherwise how could it have occurred to you to think it? And so it goes on to infinity, and this infinity is enclosed in an instant."

Bohr ultimately resolved this confusion of mind by enunciating the physical principle of complementarity which will be discussed below.** Kubie, as noted above, would possibly resolve the confusion by noting the complementary operation of a non-symbolic preconscious and a symbolic conscious.

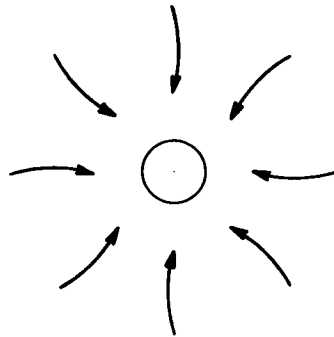
*See also footnote at end of section.

**Perhaps this precedent is sufficient justification for pursuing the line of inquiry of this section.

A reasonable objection to the above discussion is: if the preconscious is of such fundamental importance to mental operation, why has it taken mankind so long to discover it? This objection is only partially answered by observing that the analytical modes of thought, though dominant in occidental thought, have never eliminated other thought modes. A more complete answer is to observe that "western civilization" includes only a fraction of mankind and that non-symbolic thought is much more characteristic of the oriental civilizations. Undiluted oriental thought is too much for the typical occidental to take. A good description for present purposes is given by Abegg (Ref. 6.33):

"East Asian thought does not follow a straight line but consists of enveloping or encircling moves; shown diagrammatically, it appears something like this:

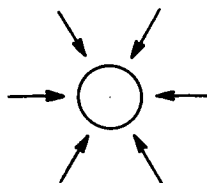
FIG. 6-21



"The East Asians at first make nothing more than small advances in no particular direction, and the arrows only then turn towards a definite centre, the aim or product of the thought, when they sense this centre. At first, therefore, there is only a cloud of arrows flying in all directions and following no pattern, until they gradually turn towards a central point. These advances are a mixture, psychically considered -- partly intellectual, partly emotional, based partly on the senses and partly on the will."

"A perfect example of such thinking is when all the arrows succeed in closing right in on the object like this:

FIG. 6-22



" In this case the object is not merely reached by the intellect or by feeling, but it is completely comprehended by the whole psyche. It is not enough, then, for only one or two arrows to hit the bull's-eye, for in such a case the object has not yet been completely attained. "

"This way of thinking possesses the advantage that one very quickly knows roughly what the issue is, but it also has the disadvantage that one rarely knows exactly what it is all about. It seldom happens that one shoots completely at random, as with intellectual thinking, but it is equally rare for all the arrows, or at least enough of them, to come near enough to the mark. "

"The chief characteristic of this way of thinking lies in the fact that it remains constantly aware of the relative value of the actions of individual functions. One function hereby controls the other; sensations control feeling, the intellect and the will, the intellect controls the sensations, the emotions the will and so on - and all are directed collectively by the physical centre, the middle point of the 'self'. "

The typical reaction of the well trained occidental engineer or scientist to this description will be negative, to say the very least. It will, in fact, be similar to the reaction to Bohr's complementary principle as described by Rosenfeld, (Ref. 6-32).

"Complementarity is no system, no doctrine with ready-made precepts. There is no via regia to it; no formal definition of it can even be found in Bohr's writings, and this worries many people. Bohr was content to teach by example. The nature of this relation he regarded as sufficiently illustrated by his analyses of the limits of validity of classical physical concepts. "

A reasonably good textbook type description of the complementarity principle is that of Schiff (Ref. 6-34). "In order to understand the implications of the (Heisenberg) uncertainty principle in more physical terms, Bohr introduced the complementarity principle in 1928. This principle states that atomic phenomena cannot be described with the completeness demanded with classical dynamics; some of the elements that complement each other to make up a complete classical description are actually mutually exclusive and these complementary elements are all necessary for the

description of various aspects of the phenomena. From the point of view of the experimenter, the complementarity principle asserts that the physical apparatus available to him has such properties that more precise measurements than those indicated by the uncertainty principle cannot be made. "

Heisenberg suggests that one can escape from this by using mathematic formalism. (Ref. 6-35). The essence of the complementarity principle is, however, that this escape is something of a delusion. The mind actually must use classical thought modes to think about quantum mechanical effects, because, in the last analysis, quantum mechanical phenomena can only be detected by macroscopic experimental effects.

The relationship between the complementarity principle and oriental thought modes is clear from the diagrams in the Abegg quotation given above. The complementarity principle asserts that in thinking about quantum mechanical phenomena, i. e. , where Planck's constant 'h' is important, there must be at least two arrows in each of the diagrams. Bohr himself realized that a connection existed between oriental thought and his complementarity principle (Ref. 6-36, p 20).

It is clear that some form of Heisenberg's uncertainty principle, and hence some form of the complementarity principle, must apply in biology and psychology because experiments necessarily perturb the system being studied. It is possible that many of the difficulties of the biologist and the psychologist stem from a constant, analogous to the Planck constant, which is relatively large.

The classic description of the creative process given by Poincare (Ref. 6-31) will now be quoted in part.

"Perhaps we ought to seek the explanation in that preliminary period of conscious work which always precedes all fruitful unconscious labor. Permit me a rough comparison. Figure the future elements of our combinations as something like the hooked atoms of Epicurus. During the complete repose of the mind, these atoms are motionless, they are, so to speak, hooked to the wall; so this complete rest may be indefinitely prolonged without the atoms meeting, and consequently without any combination between them. "

"On the other hand, during a period of apparent rest and

unconscious work, certain of them are detached from the wall and put in motion. They flash in every direction through the space (I was about to say the room) where they are enclosed, as would, for example, a swarm of gnats, or, if you prefer a more learned comparison, like the molecules of gas in the kinematic theory of gases. Then their mutual impacts may produce new combinations."

"What is the role of the preliminary conscious work? It is evidently to mobilize certain of these atoms, to unhook them from the wall and put them in swing. We think we have done no good because we have moved these elements a thousand different ways in seeking to assemble them, and have found no satisfactory aggregate. But, after this shaking up imposed upon them by our will, these atoms do not return to their primitive rest. They freely continue their dance."

"Now, our will did not choose them at random; it pursued a perfectly determined aim. The mobilized atoms are therefore not any atoms whatsoever; they are those from which we might reasonably expect the desired solution. Then the mobilized atoms undergo impacts which make them enter into combinations among themselves or with other atoms at rest which they struck against in their course. Again, I beg pardon, my comparison is very rough, but I scarcely know how otherwise to make my thought understood."

"However it may be, the only combinations that have a chance of forming are those where at least one of the elements is one of those atoms freely chosen by our will. Now, it is evidently among these that is found what I called the good combination. Perhaps this is a way of lessening the paradoxical in the original hypothesis."

"I shall make a last remark: when above I made certain personal observations, I spoke of a night of excitement when I worked in spite of myself. Such cases are frequent, and it is not necessary that the abnormal cerebral activity be caused by a physical excitant as in that I mentioned. It seems, in such cases, that one is present at his own unconscious work, made partially perceptible to the over-excited consciousness, yet without having changed its nature. Then we vaguely comprehend what distinguishes the two mechanisms, or, if you wish, the working methods of the two egos. And the psychologic observations I have been able thus to make seem to me to confirm in their general outlines the views I have given."

"Surely they have need of it, for they are and remain in spite of all very hypothetical: the interest of the questions is so great that I do not repent of having submitted them to the reader. "*

Making due allowance for the different figures of speech, this shows important fundamental similarities to Abegg's description of oriental thought. The reader is reminded that if Kubie's thesis, referred to above, is correct, then the word "preconscious" should be substituted for the word "unconscious" in Poincaré's discussion.

It should be noted that the Poincaré description places approximately equal weight on the two elements of conscious and preconscious mental activity. If the present discussion seems to emphasize the latter it is only because occidental thought tends to emphasize the former at the expense of the latter. This emphasis is particularly striking in the educational system.** (Ref. 6-29).

If a fundamental principle of the logic of the nervous system is being discussed above, the reader raised on pragmatic, technological method will immediately ask, "How can this be made into a hardware item?"

This question will be answered by discussing the problem of pattern recognition bearing in mind the combined digital and analogue operation of the brain. Pattern recognition devices fall into two general categories:

1. Given the probable existence of patterns which fall into pre-specified classes, the device must assign each pattern to its proper class.
2. Given a set of stimuli, the device must determine that patterns exist in the set and classify the patterns usefully.

Well known devices exist in both categories. The first category is the simplest and is illustrated by check sorting machines which read characters which are predesigned to be immediately legible to a human reader. The second category is best illustrated by the human brain in its efforts to respond successfully to the natural environment. Successful pattern recognition devices which have been developed in engineering applications invariably

* The present writer asks the same indulgence of his readers.

** The typical classroom description of the scientific method is simply hypothesis-analysis-conclusion.

appear to fall into the first category because prespecified classes are always involved. This is generally true of the perceptron, which is discussed below, although the perceptron concept does lead to an interesting engineering definition of second category pattern recognition problems.

To discuss the pattern recognition problem effectively it is desirable to discuss a well defined engineering problem. It is desirable that the problem have practical usefulness, be complex enough to require a fundamental approach and simple enough to offer some possibility of solution. The problem of recognition of Latin alphabet characters and Arabic numerals is just such a problem. The capabilities of the device to be useful are:

1. To recognize printed upper case and lower case Latin alphabet characters and Arabic numerals.
2. To be insensitive to differences in type characteristics.
3. To be insensitive to small changes in the positions of the characters relative to the field of view of the device.

Pahl and Johnson (Ref. 6-37) have developed an interesting device using conventional engineering techniques for identification of lower case Russian characters. The device is based on a simultaneous five fold scan of the letters as shown in Figure 6-23. Scans 1 and 5 determine upward and downward extensions of letters. Whenever a scan crosses a line a pulse is generated. A horizontal line gives a single pulse as, for example, in scan 2 across the character 'r' in Figure 6-23. Timing of pulses is used to determine letter extensions to the left and to the right. The number of pulses per scan, and presence or absence of upward and downward extensions and of left or right extremities permits classification of the characters using the categories listed in Figure 6-23. All Russian characters are also listed in Figure 6-23. The device correctly identifies 30 of 32 lower case Russian characters and uniquely determines 26 of 32 characters. The authors do not state which characters give difficulty but they claim that a seven scan system would remove the remaining ambiguities.

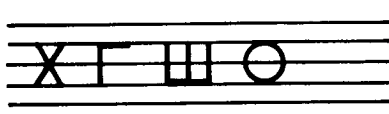
The success of the system is deceptive, which becomes apparent when it is realized that a really useful device must also be able to read upper case letters and numerals.

The introduction of numerals raises serious problems. Even a

FIG. 6-23

SCAN METHOD OF RUSSIAN CHARACTER IDENTIFICATION

Category	Characteristic	Scan Number	No. Letters Included
1	upward extending	1	3
2	downward extending	5	6
3	one pulse	2	17
4	two or three pulses	2	14
5	one pulse	3	16
6	two pulses	3	10
7	three pulses	3	6
8	one pulse	4	20
9	two pulses	4	9
10	three pulses	4	2
11	one extremity left	2, 3, 4	4
12	two extremities left	2, 3, 4	4
13	one extremity right	2, 3, 4	3
14	two extremities right	2, 3, 4	5



А а	З з	О о	Х х	Ы ы
Б б	И и Й й	П п	Ц ц	Ь ь
В в	К к	Р р	Ч ч	Э э
Г г	Л л	С с	Ш ш	Ю ю
Д д	М м	Т т	Щ щ	Я я
Е е	Н н	У у	Ъ ъ	Ө ө
Ж ж		Ф ф		У у

seven scan system of the Pahl and Johnson type would not distinguish among 1, 2, 3, 5 and 7 and the Russian characters Г, Е, З, Ж, and Т. The presence of numerals and upper case letters would thoroughly confuse category 1 in Figure 6. 23. The numerals 4, 6, 8, 9, 0 can be mutually distinguished by horizontal scanning, but confusion with letters is to be expected. It seems clear that removal of all ambiguities will require vertical as well as horizontal scanning.

If seven horizontal scans are actually required to distinguish lower case letters, and if upper case letters and numerals are almost twice the height of lower case letters, it is reasonable to assume that at least thirteen horizontal scans will be required. Since the width of letters tends to be less than the height, and since the horizontal scans come close to successful identification, it is reasonable to assume that five vertical scans would be adequate.

If thirteen horizontal scans are required it is apparent that precise positioning of the image is important. A vertical shift of half the distance between scans would completely upset the accuracy of identification. Vertical positioning must be accurate to better than 4% of the height of upper case letters. Similarly, horizontal positioning, assuming five vertical scans, must be accurate to within 10% of the width of a typical letter. It is possible in some applications to insist upon accurate positioning but in device applications such as reading machines for the blind and language translation machines this will probably prove impractical. Another approach is to have parallel logic circuitry connected to the several scans such that one set of logical circuitry is correct for any possible image position. If the allowed position error is any combination of a positive or negative displacement of one scan either vertically or horizontally, then nine parallel sets of logical circuitry would be required with additional circuitry to decide which set is correct. If position errors of as much as two scans are allowed, then 25 parallel sets of logic circuitry are required. This corresponds to errors of $\pm 20\%$ in height and $\pm 50\%$ in width. This approach also would require additional scans: four more horizontal scans and four more vertical scans in the case of allowed two scan position errors.

It is clear that the Pahl and Johnson device, which at first appears

close to a satisfactory solution to the problem, would require much greater complexity to yield a practical solution.

This example illustrates several general principles in pattern recognition. The first general principle is that in pattern recognition the number of bits of information processed is much greater than the bits of information gained in the selection. In the Pahl and Johnson device correct selection corresponds to an information gain of less than five bits although there are 14 information categories used corresponding to 14 bits of information which must be processed. It is interesting to investigate relationships between the information gained in a pattern recognition decision and the amount of information which must be processed to make the decision. This can be done, for example, by considering the recognition of upper case letters by pulse counting in single scans. The ground rules are: (1) the scans may be horizontal or vertical, (2) the letters can be placed such that the scans do not pass through unusual parts of the letter, i. e., a horizontal scan does not pass through the horizontal bar in A, (3) a scan must not rely on passing through an unusual part of a letter. In Figure 6.24, it is clear that A and B can be distinguished by a single vertical scan. The choice increases information by one bit and the information which must be processed is one bit. Letter C is then added and a second scan is required. As additional letters are added, either an additional scan is required or an additional decision on an existing scan is required. Assuming equal probabilities, the information gained in selection of a letter, I_s , is given by:

$$I_s = 1.44 \ln N$$

where N is the number of letters and the logarithm is to the base e . The information which must be processed, I_p , is given by:

$$I_p = \sum_i n_i \ln n_i$$

where n_i is the number of choices in scan i and the summation is over all scans. Figure 6-24 illustrates the scans required to distinguish among different sets of letters, gives the choices possible for each scan and gives I_s and I_p for each combination of letters.

In Figure 6-25, I_p is plotted versus I_s on a semilog plot. The five scan system of Pahl and Johnson requires 14 categories to identify 26

CHARACTER GROUP	PULSES PER SCAN					INFORMATION PROCESSED	INFORMATION GAINED
	1	2	3	4	5		
A B	2,3					1	1
A B C ²	2,3	1,2				2	1.59
A B C D ²	2,3	1,2	1,2,3			3.59	2.32
A B C D E ²	2,3	1,2	1,2,3			3.59	2.32
A B C D E F ²	2,3	1,2	1,2,3	1,2		4.59	2.58
A B C D E F G ²	2,3	1,2	1,2,3	1,2		5.18	2.81

FIG. 6-24

SCAN IDENTIFICATION OF CHARACTERS

CHARACTER GROUP	PULSES PER SCAN					INFORMATION PROCESSED	INFORMATION GAINED
	1	2	3	4	5		
	1, 2, 3	1, 2	0, 1, 2, 3	1, 2		6.59	3
	1, 2, 3	1, 2	0, 1, 2, 3	1, 2		6.59	3.16
	1, 2, 3	1, 2	0, 1, 2, 3	1, 2		6.59	3.31
	1, 2, 3	1, 2	0, 1, 2, 3	1, 2		6.59	3.46

FIG. 6-24 (CONT'D)

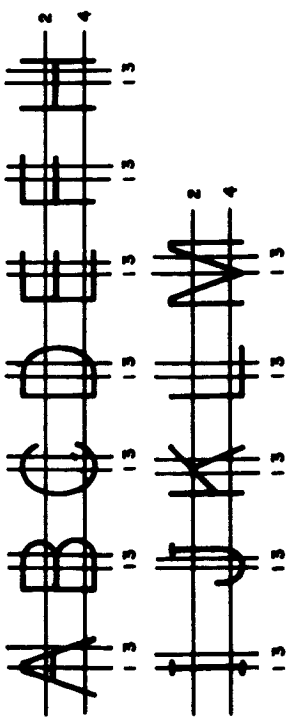
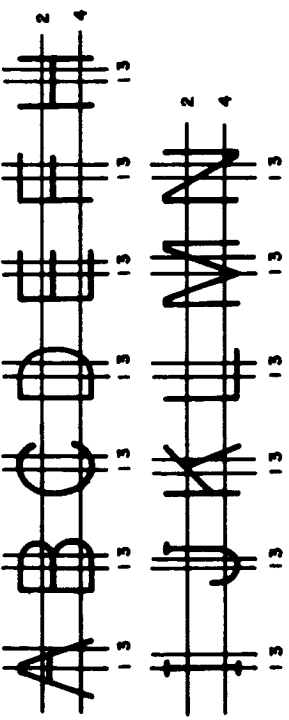
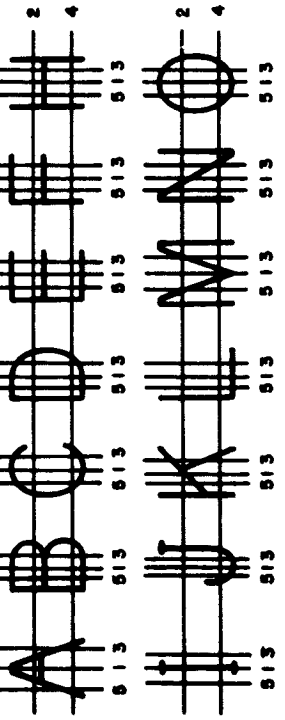
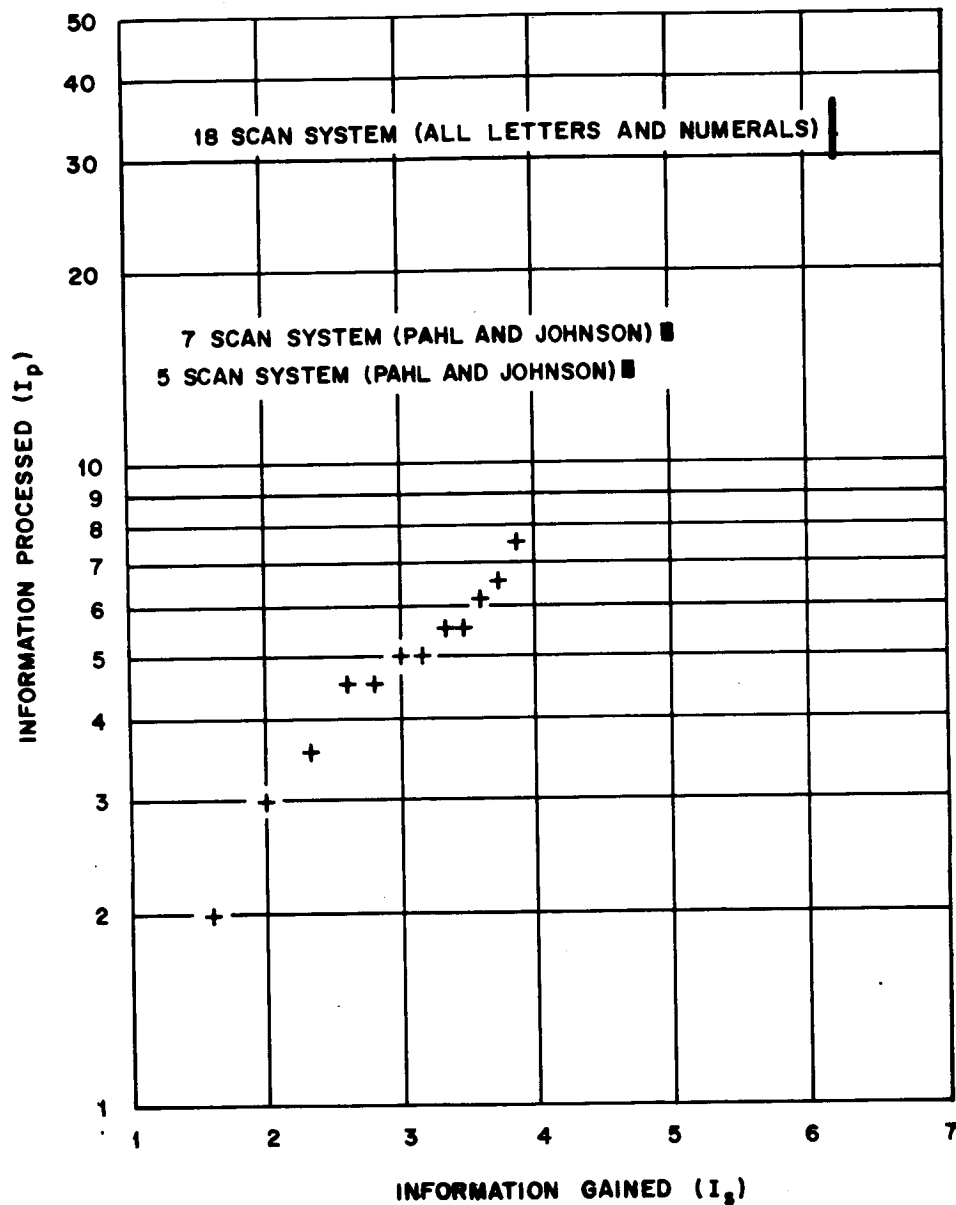
CHARACTER GROUP	PULSES PER SCAN					INFORMATION PROCESSED	INFORMATION GAINED
	1	2	3	4	5		
	1, 2, 3	1, 2, 4	0, 1, 2, 3	1, 2, 4		6.77	3.57
	1, 2, 3	1, 2, 3, 4	0, 1, 2, 3	1, 2, 4		6.77	3.69
	1, 2, 3	1, 2, 3, 4	0, 1, 2, 3	1, 2	0, 1, 2, 3	8.59	3.80

FIG. 6-24 (CONT'D)

FIG. 6-25

INFORMATION PROCESSED, I_p , VERSUS INFORMATION GAINED, I_s , IN SCAN RECOGNITION SYSTEM



characters uniquely, or 14 bits of information must be processed to make a 4.5 bit decision. This is shown in the plot. They estimated that two additional scans would achieve unique identification of the 32 lower case characters. Assuming an additional bit for each additional scan, this point is also plotted. It was estimated above that 18 scans would be required for recognition of the 64 upper and lower case Russian characters and the digits. If $n_i = 4$ for each scan then $I_p = 36$ and if $n_i = 3$ for all i then $I_p = 28.6$. It appears that the points approach a linear relationship corresponding to an equation of the form:

$$I_p = K e^{\alpha I_s}$$

where K and α are constants. From the slope of Figure 6-25, $\alpha \approx 0.65$ and also $K \approx 0.62$.

It has already been observed that addition of a letter in Figure 6-24 required either addition of a scan or addition of another possible choice in an existing scan. The amount of additional information which must be processed depends critically on which is required. If a scan is added, a whole bit of additional information must be processed. If a scan selection is changed from three choices to four choices the bit increment is 0.41. This increment is smaller for a larger number of choices in a scan. Since the selection of vertical and horizontal scans was arbitrary, this raises the question of whether a better engineering solution can be achieved. If a large number of choices per scan reduces the amount of information which must be processed, can an optimum system be designed by going to the extreme case? Suppose that the total line length of letters is measured. It is probable that line lengths could be used to identify letters. For present purposes it will be assumed that this would give unique identification. One way of measuring line length might be to count the number of photodetectors in a retinal array which are triggered. It is clear that such a solution has a serious difficulty. To select among, say 62 characters, measurement accuracies of less than 1% would be required. Such a system would also be very subject to noise. There is, however, an even more fundamental objection. Line length is a very detailed property of a character. The device would be helpless in dealing with different kinds of type. Line length has one great advantage: it is independent of position. Generally speaking these properties are characteristic of any device which relies on a single scan.

It is possible to go to the other extreme and use as many scans as there are characters, and design each scan to have a choice between two possibilities. This would minimize the information processing required but the positional relationship of the scans to the letters would be very important and the device would be very sensitive to changes in position and kinds of type, and to noise.

There appears to be a dilemma. If one designs to minimize information processing required, the resulting device is very sensitive to noise, type of print and possibly position. If, on the other hand, one designs a device to be relatively insensitive to these perturbing conditions, the information processing requirement increases - and the increase becomes very great for increasing numbers of patterns which are to be distinguished.

It is reasonable to turn to bionics for insight into the problem.

The perceptron is certainly a product of bionics. In the words of Block, "The perceptron is a self organizing or adaptive system proposed by Rosenblatt. Its primary purpose is to shed light on the problem of explaining brain structure. It also has technological applications as a pattern recognizing device.... The technological aspects are not completely irrelevant, however, since a model, no matter how appealing it may appear from the point of view of structural similarity, must also be judged on the basis of its performance." (Ref. 6-38).

There are various types of perceptrons but the photoperceptron is typical. Photoperceptrons have been built and successfully operated in pattern recognition, but there seems to be no engineering problem which has been solved using a perceptron.

The perceptron constructed by Block consists of a "retina" composed of a 20 x 20 matrix of photocells, 512 associator units, and eight response units. Each photocell is connected to up to forty associator units. Whether or not a particular associator unit is connected to a particular retinal photocell is determined randomly. The associator units are in turn connected to the response units in a random, many to one, connection. The connections from the photocells to the associator units may tend to trigger or may tend to inhibit activation of the associator units. The associator units have the property that each time they operate a functional characteristic is changed.

When an image is placed on the retina the response units register in a particular way. How the response units react to a single particular

image is random. The change in the association units tend to make a repetition of the image lead to the same response. The perceptron can, therefore, learn. If exposure to an image leads to an incorrect response there is provision for cancelling the effect of that exposure. That is, mistakes can be unlearned.

Rosenblatt (Ref. 6. 27) developed a theory that predicted important features of perceptron behavior. Perceptrons have been constructed and successfully operated. There is reason to believe that perceptron behavior does in certain important ways model brain behavior.

The perceptron can be trained to separate a set of different images into arbitrary classifications. As the number of classifications is increased, however, the perceptron's ability to do this decreases. More to the point, if a set of images which are in several classes, such that there are similarities within each class, are flashed on the retina, the perceptron can learn to distinguish between these classes. If a new image is placed on the retina, the perceptron will classify the image. The accuracy of classification improves with additional exposure - a result just the opposite to that for arbitrary classification mentioned above.

The association units in the perceptron are fairly complicated electrical circuits. To make engineering applications it is necessary to know the minimum number of association units required to obtain perceptron behavior. Rosenblatt in his original report showed that the probability of an association unit being triggered when an arbitrary image is placed on the retina is an important parameter in predicting the behavior of the perceptron. This probability is a rapidly varying function of the number of retinal points if the number of retinal points is less than about twenty. For more than twenty retinal points this probability is almost independent of the number of retinal points.

Block performed an interesting experiment using a laboratory perceptron having 240 association units. He trained the device to distinguish between the letters E and X. He then successively removed 1/2, 3/4 and 7/8 of the association units. The experiment was performed both when a letter was the only image on the retina and also when "noise" in the form of random dots was projected with the letter into the retina. The results of the experiment are shown in Figure 6. 26. Even with only thirty association units remaining the perceptron was moderately effective. Evidently most perceptrons

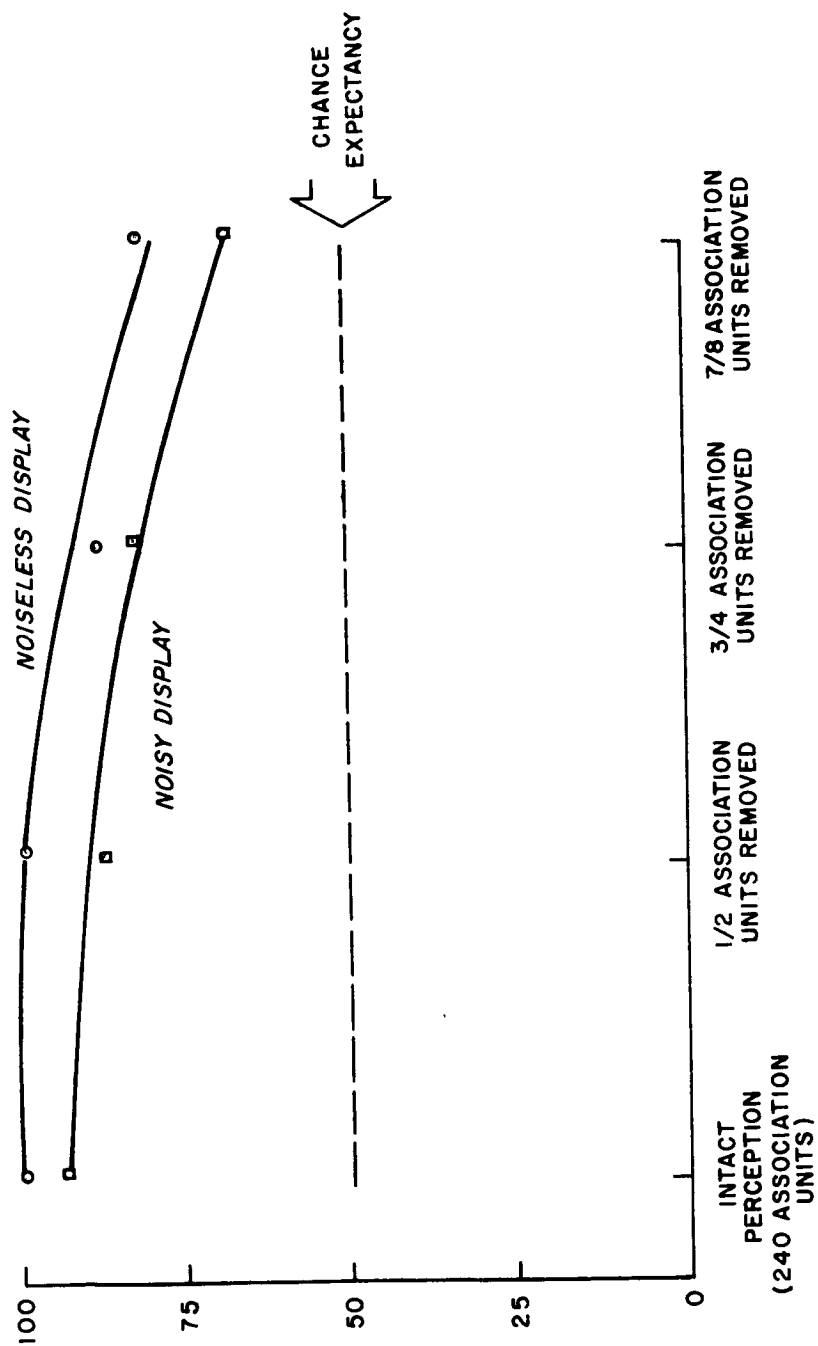


FIG. 6-26

EFFECT OF ASSOCIATION UNIT REMOVAL ON TRAINED "E" - "X" DISCRIMINATION

(REF. 6-38)

which have been designed and/or constructed are oversized from the standpoint of the solution of useful engineering problems. It will be seen below that the design of minimum performance perceptrons might be extremely useful.

Let us consider again the letter recognition problem. It has already been agreed that the use of single scan or many scan methods have inherent difficulties from the standpoint of sensitivity to noise and position. Use of an intermediate number of scans requires the processing of much more information than the pattern selection actually represents. The earlier discussion of certain psychological phenomena suggested a duality in mental operations: digital vs. analogue, conscious vs. preconscious, precise vs. probabilistic. One side of this dualism suggested the development of the perceptron. The combination of these ideas suggests the question: does the minimal engineering solution to the pattern recognition problem require the use of a combination of digital logic and perceptron logic? It has been noted by others that perceptron performance can be improved for special applications by incorporation of nonrandom connections. The question just advanced is much stronger. It asks whether the optimum engineering design for a reasonably difficult pattern recognition problem necessarily involves a combination of digital and random devices.

The importance of the question is sufficient to propose the following theorem whose truth has yet to be established:

"The Fundamental Theorem of Nervous System Organization states that in a pattern recognition problem whose complexity exceeds a certain threshold complexity, the optimum engineering device for solution of the problem includes both digital logic and perceptron logic units."

The fundamental theorem could be true for pattern recognition problems of either the first or second category or for both categories (see page 6-62).

It is here submitted that the letter recognition problem, aside from its practical usefulness, is appropriate for an investigation of the fundamental theorem. Examination of Figure 6.24 will show that addition of a scan for distinguishing an additional letter is almost always required by ambiguity with only one other letter. The additional information provided

by the new scan is entirely redundant for most distinctions among letters. This suggests the design of a system having the following characteristics:

1. A small digital system with a small number of scans with the scans designed to reduce ambiguities to comparisons between two or three characters.
2. A minimum performance perceptron to remove remaining ambiguities.

Current design perceptrons could easily be adapted to this approach. The horizontal scans of Pahl and Johnson would correspond to counting the number of activated photocells in a row of photocells of the retina.* The retina approach would of course, provide a greater flexibility in scan selection. The line length scan discussed above would correspond simply to counting the number of photocells activated in the whole retina. The photocell outputs for selected scan sets of photocells can be digitally treated using the Pahl and Johnson techniques. The perceptron itself can be made into a minimal performance perceptron by disconnecting association units.

If a composite digital and perceptron system is actually optimum it may be necessary to select the proper scans for digital analysis. One criterion for selection is that the ambiguous letters should be as unlike as possible to maximize the effectiveness of the minimum performance perceptron.

Combining the digital output information and the perceptron output information in an optimum way is also an interesting problem.

It seems clear that properly motivated effort within the state of the art should be able to establish whether the optimum pattern recognition system for a given set of characters actually requires combined digital and perceptron techniques. If an affirmative result is obtained, it is reasonable to expect that it will then be possible to decide the generality of this assumption.

The discussion of pattern recognition problems has this far been restricted to first category pattern recognition - pattern recognition in which the patterns are assigned to pre-specified classes. A second category of pattern recognition problems are those in which the device determines whether or not a set of stimuli contains patterns which permit useful classification. It is also been noted that if the patterns taught to the perceptron have no

* See "Note added in proof" at end of section.

inherent similarities, the capacity of a perceptron to classify new patterns decreases as more patterns are learned. In Rosenblatt's words, "In such an 'ideal environment' the probability of a correct response diminishes towards its original random level as the number of stimuli learned increases." He adds that, "In a differentiated environment, where members of a stimulus class associated to a given response are correlated in some respect. , the probability of a correct repetition of a previously reinforced response will not decline to the original random level, no matter how many associations of each class are imposed on the system. In the differentiated environment, the probability of correct response to any stimulus of one of the learned classes increases with experience. , regardless of whether or not the particular stimulus has ever been seen before" (Ref. 6-27). Reversal of the logical order of Rosenblatt's statements leads to the conclusion that the perceptron can, in principle, solve second category pattern recognition problems. Suppose one has a large number of stimuli which can be separated into two classes on the expectation that there exists a systematic but unknown difference between the classes, that is, there is a pattern in either or both of the classes. If there is such a pattern difference, then the perceptron will detect this by continuing to make correct classifications for additional stimuli. This is second category pattern recognition.

Rosenblatt hastens to add in Reference 6-27 that the similarity of forms, on the basis of which generalized recognition is possible, is not an intrinsic character only of the stimuli but is relative to the structure of the perceiving system. He seems to imply, without explicitly stating, that classification thus made by a perceptron may not be useful to a human observer. The implicit nature of his statements is unfortunate, because when the question is explicitly considered it becomes clear that relativity to the particular perceptron is not necessarily a handicap in useful application. Suppose a perceptron performs a second category pattern recognition. If, then, the perceptron properties are changed in a small but non-random fashion and the second category pattern recognition disappears then not only is the existence of a pattern established but also something of its properties. This is approaching remarkably closely to the creative thought modes of the brain. The non-random variation of the perceptron properties shows, in fact, a remarkable similarity to the scanning properties of the brain which have been discussed

by Wiener (Ref. 6-30).

More recently Rosenblatt has shown greater interest in second category pattern recognition problems. He has indeed expanded his theory in an extremely fascinating way to show how a perceptron could be modified to handle second category pattern recognition (Ref. 6-39). The essential point of his newer treatment is that the association unit is designed such that the impulses leaving the unit still obey the all-or-nothing principle but that the frequency of the pulses varies continuously with the strength of the input to association unit. The similarity of this to the frequency modulation of neural impulses discovered by Adrian is apparent. Rosenblatt calls such association units, "continuous transducer neurons."

Rosenblatt has made an interesting observation. "One important difference between the learning curves observed for spontaneously organizing perceptrons and those of humans, however, is worth noting: The learning curves are convex, whereas those for a human subject under similar circumstances would certainly be concave. That is, if a perceptron is required to distinguish horizontal from vertical bars, in a spontaneous learning experiment, it quickly learns to classify the first sixty or seventy percent of the cases and takes longer and longer to establish the correct response for the few stimuli still unclassified. A human subject, in such a problem, once he has established an 'insight' would undoubtedly jump to 100% accuracy immediately thereafter."

The reason for this difference can, perhaps, be guessed if one considers what an 'insight' is. A person generally says he has had an insight when he can verbally define some property of the phenomenon he is investigating. Verbal definition is, however, non-random. Its logic is digital in character and it is not surprising that the perceptron will not show insight. Again, there is evidence that optimum pattern recognition requires both perceptron and digital techniques. Non-random variation of perceptron parameters in a scanning process has already been suggested as one way of incorporating non-random behavior into a perceptron.

The discussion of this section can be summarized in the following statements.

1. Examination of mental operations from a psychological and a philosophical viewpoint suggests an inherent duality which can be variously

expressed as conscious/preconscious, digital/analogue or as demonstrating a complementarity principle.

2. Examination of the specific pattern recognition problem of alphabetic character identification suggests that this duality can be incorporated into hardware by a device using both digital logical processing and perceptron techniques.

3. The Fundamental Theorem of Nervous System Organization has been proposed. This states that an optimum pattern recognition system necessarily involves both digital and random logical techniques.

4. The alphabetic character identification problem has been suggested as offering an ideal opportunity for investigation of the fundamental theorem.

5. The same investigation would also result in a solution of an engineering problem of great practical significance.

6. Pattern recognition problems of the second kind, i. e., problems in which there are no predesignated categories, have been discussed. It has been suggested that one difference between perceptron behavior and human behavior may be attributable to the absence of non-random or digital logic in the perceptron.

It is desirable to conclude this section with the observation that there is a vast literature on what has here been designated as the logic of the nervous system. This vast literature creates what may be called a "climate of opinion." The observation that the brain shows a remarkable duality in its operation is part of the climate of opinion, although this is not always considered.* Examination of specific pattern recognition problems against this background has led to the statement, without proof, of the Fundamental Theorem of Nervous System Organization. This states that in a pattern recognition problem whose complexity exceeds a certain threshold complexity, the optimum engineering device for solution of the problem includes both digital logic and perceptron logic units. It is suggested that practical engineering investigation of this theorem would yield valuable practical results.

* Note added in proof: S. B. Ruth has recently discussed hybrid computers from an engineering viewpoint (Ref. 6-40). D. H. Hubel has shown the presence of interconnection among retinal elements similar to the vertical and horizontal line detector connections discussed above. (Ref. 6-41).

Note: Many optical character recognition systems are described in a recently published book (Ref. 6-53). The RCA multi font reading machine described by W. J. Hannan uses "mask matching" which is logically equivalent to the "one scan for each character" method discussed above. W. S. Homes, et al, describe first efforts in use of the perceptron explicitly for character recognition. A variety of multiple scanning systems are described by other authors. The concept described in this section as the "fundamental theorem of the nervous system" is not discussed although there are vague hints. Information in this valuable book does not, therefore, require serious modification of the present discussion.

6.3 Applications

6.3.1. Neuristor

A neuristor is a signal propagation device which operates in a manner more or less similar to that of the nerve axon - an active transmission line. The essential characteristic is that the signal is transmitted without attenuation by deriving its energy from energy stored along the transmission line.

It is impossible to study biosensors without becoming aware of the significance of the system of neural conduction, by which information is taken from the sensor and transmitted to the brain. A strong impression is gained that if the mechanism of nerve conduction were completely understood, then it would be much easier to understand the various sensory mechanisms.

From the point of view of bionic research, the neural method of information conduction would appear to be of the greatest significance. It differs not only in detail, but also quite fundamentally, from the methods of information conduction normally employed in man's technology. Information transmission in nerves is an ACTIVE process, whereas our own methods are generally PASSIVE. Copper wires and cable prove now to be a perfectly satisfactory medium for information transmission in some applications; active transmission may be necessary for particular future application. A number of neuristors using electronic components have been investigated, and there is an excellent review of the subject available (Ref. 6-43).

It seems probable that the value of neuristors will result from availability of neuristors having effectively small diameters. Nature achieves this objective by using materials with unusual non-linear properties. In this section two new concepts for possible small diameter neuristors will be discussed. The information transmission capacities of neuristors will then be investigated in an effort to define the properties of a useful neuristor. It must be constantly remembered that effective ways of information transmission are already available and neuristors must demonstrate relative advantages in particular applications to justify their use.

Two neuristor concepts relying on material properties take advantage of a cryogenic effect, superconductivity, and of the Hall effect as described below. These were conceived at Allied Research in the course of the work being reported here.

Cryogenic Neuristor

A strip of superconducting material (SM) is fixed between two copper strips (Fig. 6-27a). A magnetic field is applied to the SM. The temperature is adjusted so that the SM is just not superconducting. Under these conditions a slight reduction in the magnetic field will be sufficient to cause transition to the superconducting state. A DC potential difference is maintained between the copper strips. This will cause a small current to pass at all points along the non-superconducting SM.

Suppose that at some small local point on the SM the external magnetic field is reduced causing local transition to the superconducting state (Fig. 6-27b). The current flux through the SM (between the copper strips) will increase at this point. The magnetic field, H_L , due to this local increase in current will act in a direction to reinforce the magnetic field on one side of the local region and to cancel the magnetic field on the other side.

The effect will be to cause a transition to the superconducting state at a point adjacent to the original local region of superconductivity. The increased magnetic field due to the extra current flow at this new region of superconductivity will act to destroy the superconductivity state in the original local area. It will also act to cause transition to superconductivity at a new point further away again from the original local disturbance in the magnetic field.

It can be seen that the effect of the original local disturbance in the magnetic field will be to cause an impulse, that is, a wave of superconductivity to propagate down the line, followed by a region of relatively large magnetic field - which is analogous to the inhibited state of a nerve fiber immediately after conduction.

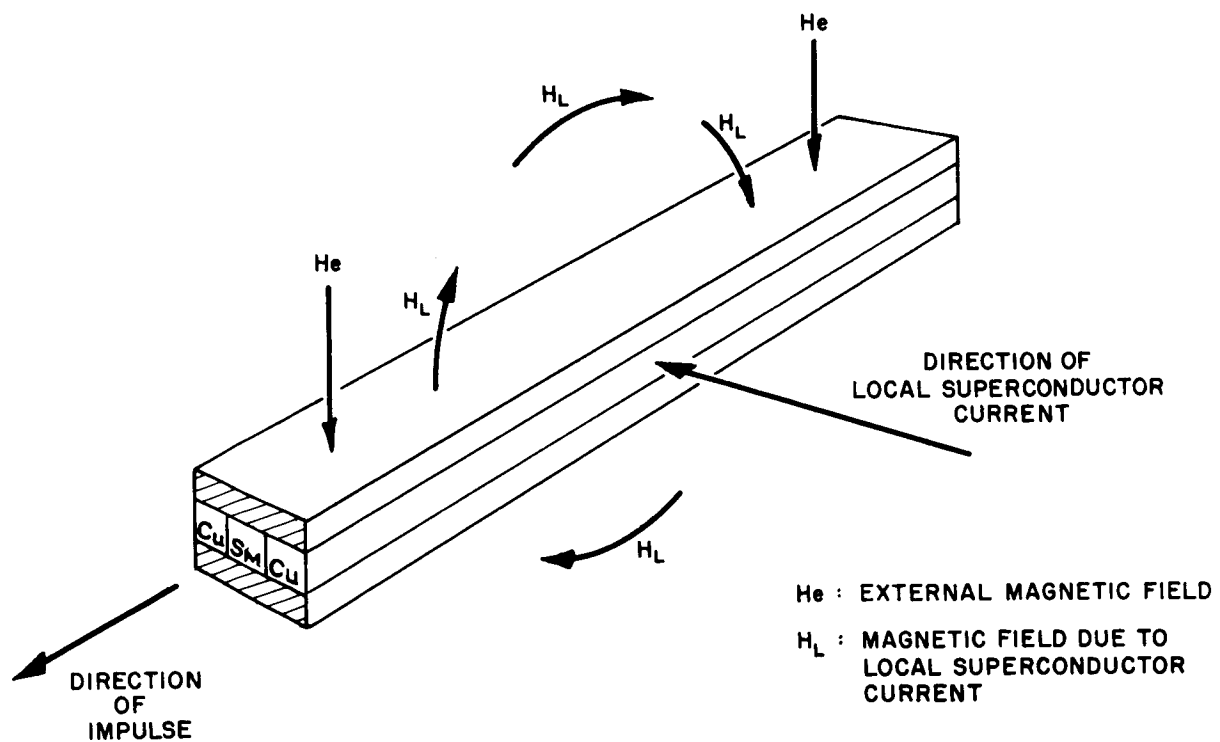
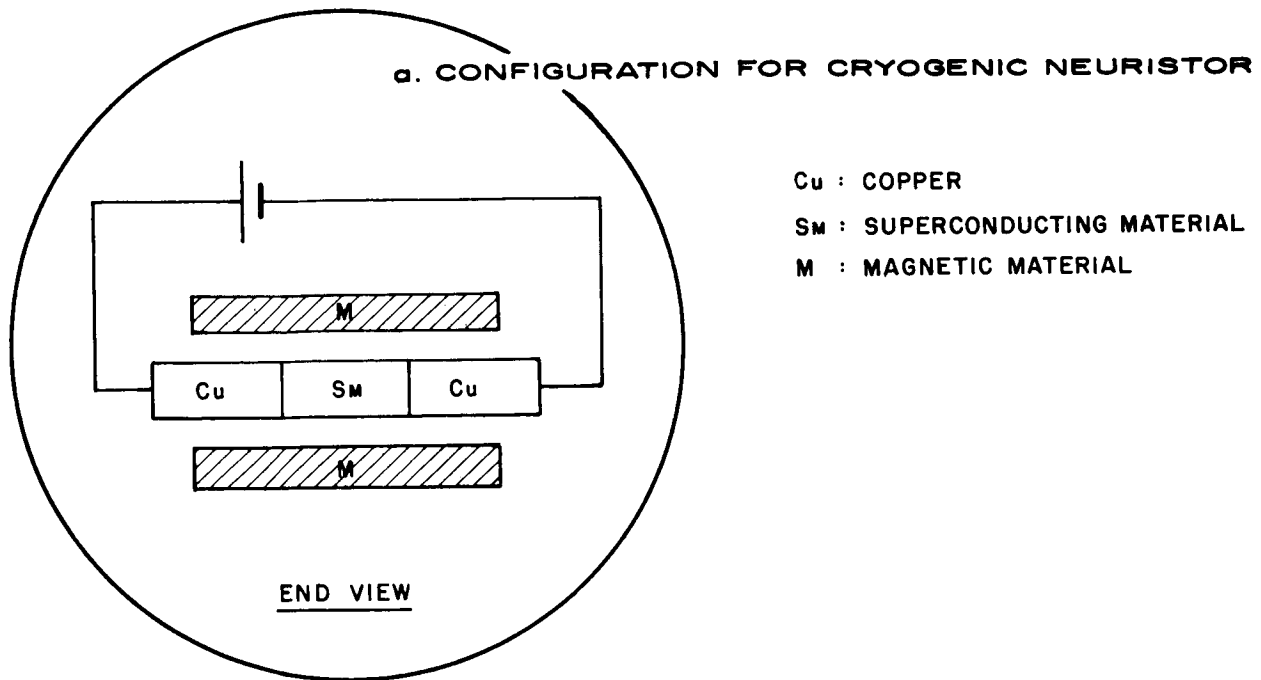
Hall Effect Neuristor

The mechanism is very similar to the one previously described. The center (core) of the line is made of Hall effect* material. It is sandwiched between semiconductor materials and copper strip as shown in Fig. 6-28a. The current between the two copper strips is non-linear due to the negative resistance characteristic of the semiconductor material. The Hall material is in a magnetic field and has a resistance adequate to nullify

*Change in resistance with change in magnetic field.

FIG. 6-27

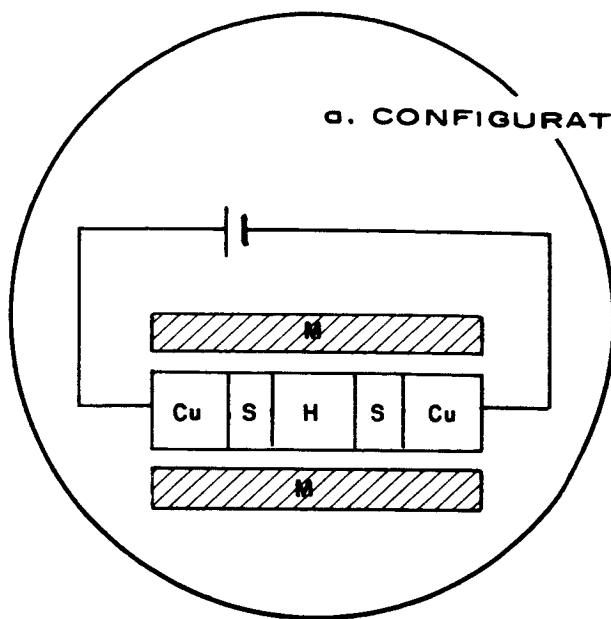
CRYOGENIC SUPERCONDUCTING NEURISTOR



b. OPERATION FOR CRYOGENIC NEURISTOR

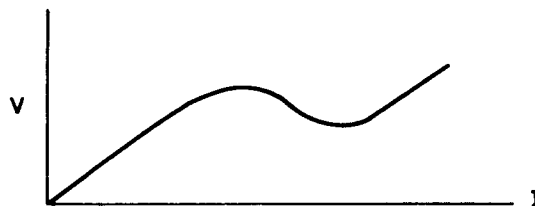
HALL EFFECT NEURISTOR

a. CONFIGURATION FOR HALL EFFECT NEURISTOR

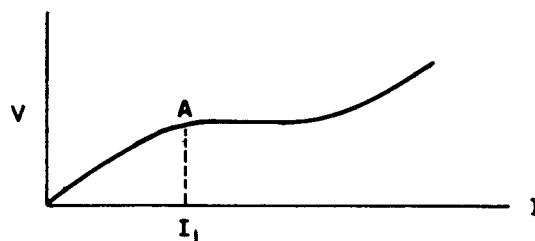


Cu : COPPER
S : SEMICONDUCTOR MATERIAL
H : HALL EFFECT MATERIAL
M : MAGNETIC MATERIAL

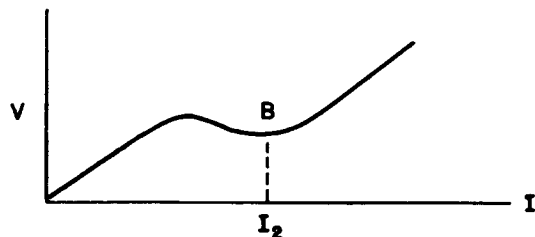
WITHOUT
HALL
RESISTANCE



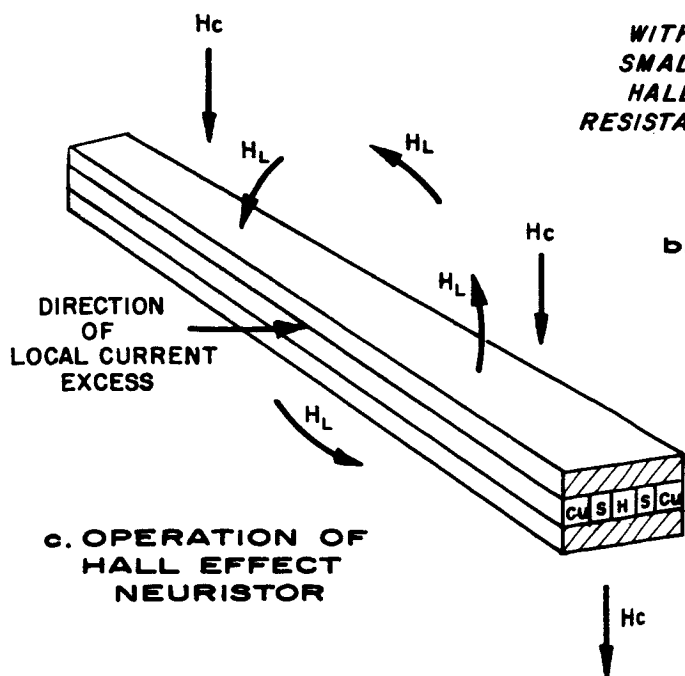
WITH
LARGE
HALL
RESISTANCE



WITH
SMALL
HALL
RESISTANCE



b. CHARACTERISTIC OF NON-LINEAR SEMICONDUCTOR CIRCUIT



c. OPERATION OF HALL EFFECT NEURISTOR

H_e : EXTERNAL MAGNETIC FIELD
 H_L : MAGNETIC FIELD DUE TO LOCAL CURRENT EXCESS

DIRECTION OF IMPULSE

the negative resistance of the semiconductor. However, a slight reduction in the magnetic field will cause a decrease in the Hall resistance adequate to give an overall negative resistance effect between the copper strips (Fig. 6-28b).

To propagate a wave, the magnetic field is locally decreased. The current between the copper strips locally increases from I_1 to I_2 (Fig. 6-28b). This increased current will cause a drop in the Hall resistance ahead of this original point (Fig. 6-28c).

A wave of increased current will propagate down this line in a manner analogous to that of the cryogenic line.

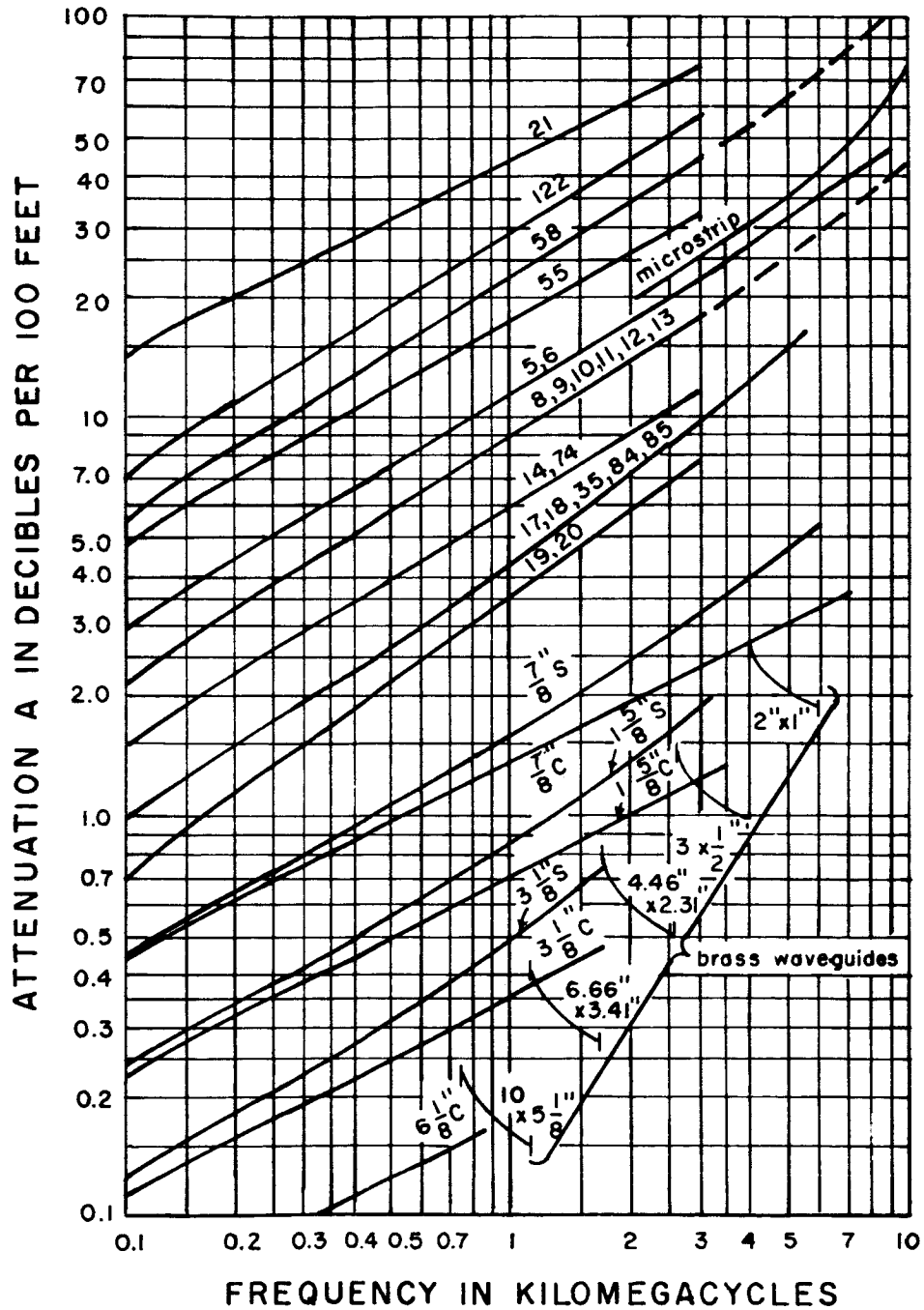
The cryogenic and Hall effect neuristors have two properties in common with biological nerve fibers. First, they can be individually small in diameter and the neuristors can be assembled into bundles. The requirement of maintaining an external magnetic field is significant disadvantage for long distance transmission but may not be a serious handicap in some applications, such as data handling equipment. Second, they both depend on non-linear properties of materials. It is possible to obtain materials in small diameter and in quantity by such processes as extrusion and deposition. Most neuristors which have been proposed are electronic devices using active and passive electronic components. Small diameter neuristors of this type are difficult to visualize.

The value of neuristors in practical engineering will be to transport information with minimum power over a distance specified by a given requirement. It is of interest, therefore, to compare the maximum rates of information transmission over nerve bundles and over cables of comparable diameter. There are numerous engineering difficulties involved in the use of neuristor bundles. It is reasonable, therefore, to insist that the information transmission rate of neuristor bundles be at least the same order of magnitude as that of comparable diameter cables.

To make the comparison between nerve bundles and cables we will first investigate the properties of coaxial cables. These are shown in Fig. 6-29 which is plots of attenuation per 100 ft. as function of frequency for cables of a variety of types (Ref. 6-44). Of particular interest are the rigid copper coaxial cables labeled 7/8"C, 1.5/8"C and 3.1/8"C. Easily commercially available styroflex cables, denoted by 7/8"S, 1.5/8"S and

FIG. 6-29

MATCHED LINE ATTENUATION IN CABLES



301

3. 1/8"S, are similar to the rigid copper coaxial cables in behavior, especially at lower frequencies. Curves of this nature may be expected to exist for cable diameters such that the skin depth of the electromagnetic radiation is appreciably less than the thickness of conductors in the cable.

One outstanding characteristic of all curves for cables is the strong linear dependence. If the attenuation is denoted by β then

$$\ln \beta = \ln A + B \ln \nu \quad (6-32)$$

where A and B are parameters.

Examination of the curves for rigid copper coaxial cables shows that the intercept A is a function of cable diameter. If natural logarithms are used, ν is in kilomegacycles and β is in decibels per 100 ft, $B = 0.5$ for the rigid copper coaxial cables.

Equation (6-32) can be rearranged to yield

$$\beta = A \nu^B \quad (6-33)$$

If nA is evaluated from the curves, A is found to vary inversely as the diameter. Equation 6-33 then becomes

$$\beta = \frac{3}{D} \nu^{0.5} \quad (6-34)$$

where D is in cm.

The cable length can be introduced by converting 100 ft. into 3×10^3 cm. Equation 6-34 becomes

$$\beta = 10^{-3} \frac{L}{D} \nu^{0.5} \quad (6-35)$$

By definition,

$$\beta = 10 \log \frac{P}{P_0} \quad (6-36)$$

where P is the attenuated power and P_0 is the unattenuated power. Equations 6-35 and 6-36 combine to yield

$$P/P_0 = e^{-(2.3 \times 10^{-4}) L \nu^{1/2} / D} \quad (6-37)$$

We will arbitrarily assume a signal of a particular frequency can reliably deliver information only if its power at the receiving end exceeds the noise at the receiving end by a factor of the order of two. The noise includes at least thermal noise. The power at the receiving end is determined by the power sent and by the attenuation. In practice, severe attenuation is not desirable because power is wasted and thermal noise is increased. It seems reasonable to require that P/P_0 exceed e^{-1} .

Consider a cable whose diameter is 0.3 cm and whose

length is 100 cm. Then $P/P_0 > e^{-1}$ if $\nu < 169$ kilomegacycles. Such a bandwidth can transmit 2.38×10^{11} bits per second assuming symmetric double signal levels.

Consider now a nerve bundle of equal diameter. The optic nerve, for example, contains 1.25×10^6 fibres in a diameter of 0.3 cm. If the absolute refractory period is the order of 2 milliseconds, then each fibre can transmit up to 4×10^2 pulses per second which corresponds to 4×10^2 bits of information per second. The nerve bundle has a maximum information capacity of 6×10^8 pulses per second. The coaxial cable can transmit 400 times as much information as the nerve fibre bundle over a distance of 100 cm.

It should be noted that this discrepancy is in part a result of materials quality. The coaxial cable contains good conduction whereas the nerve fibre contains ionic solution conductors. The difference in conductivities can be the order of 10^3 . It is not unreasonable to expect, therefore, that the impulse velocity along a nerve fibre constructed of good conducting materials, i. e., along neuristors, such as those discussed above, will be considerably greater. If the absolute refractory period were correspondingly shorter, it is reasonable to expect that the information rates of the neuristor bundle and the coaxial cable would be comparable.

From Equation 6-37, it is apparent that as D decreases, maximum ν (and, therefore, the information rate) must also decrease to maintain the same allowable attenuation. The information rate decreases with the square of the diameter. But the information rate of the nerve fibre bundle also decreases with the square of the diameter because the number of fibres in the bundle is proportional to the area. It appears, therefore, that, given comparable materials, neither the neuristor bundle nor the coaxial cable has any inherent advantage in information rate for any diameter.

The functional dependence for small cables in which skin depth is equal to or greater than cross section dimensions of the cable will be similar to that for the larger cables because attenuation is determined by area which depends on the square of linear dimensions.

It is interesting to investigate the dependence of information rate on nerve fibre diameter of a nerve fibre bundle with given total diameter.

The impulse velocity of a nerve fibre increases linearly with the diameter, but the area increases with the square of the diameter. The information rate of a bundle of given diameter will vary inversely with fibre size. The presence of large fibres is basically inefficient in rate of information transmission. The presence of large diameter nerve fibres possibly indicates the need for rapid transmission of specific kinds of information.

It is safe to conclude from the above discussion that engineering choices between neuristor bundles and conventional cables will be determined by considerations other than gross information rate capability.

Neuristors can be divided into three categories:

1. Chemical neuristors rely on a chemical reaction in a membrane. Biological nerve fibres are an example of this type.
2. Material Properties neuristors rely on a non-linear characteristic of a material. The cryogenic and Hall effect neuristors discussed above are examples of this type. Thin film neuristors using semiconductor materials (Ref. 6-43) would also be in this category.
3. Electronic Component neuristors rely on discrete electric components to produce the non-linear effect necessary for neuristor action.

The reason for introducing the above classification is that it emphasizes that production of small diameter neuristors is conceivable if the neuristors belong either in the chemical or the material property category. A diameter of ten microns is not unusual, for example, for optical fibres and these are two component systems consisting of a sheath and of a core having different indices of refraction. It is difficult to visualize small diameter electronic component neuristors.

Since neuristors will not necessarily compete with conventional methods of information transmission in rate of transmission, engineering applications must rely on the unique properties of neuristors.

One unique property of neuristors is that transit times of particular bits of information are relatively long. Neuristors could, there-

fore, act as delay lines. This would be a particularly interesting possibility if neuristors could be designed as pulse forming systems or as distortionless systems in which the shape of the output pulse is determined by the shape of the input pulse.

Arbitrarily complex digital-logic systems can be synthesized with relays. Crane (Ref. 6-43) shows that neuristors can be used to synthesize direct analogues of relay networks and that neuristors have, therefore, complete logic capability.

A major difficulty in miniaturization using conventional electronic devices is the problem of attaching connections. This would also be a problem in using bundles of small dimension neuristors. The difficulty is increased in digital logic networks where precise connections are necessary. It may well be that successful achievement of digital computers having large logic capability in small volume may be limited by the connection problem. In perceptrons, however, component precision is not required. It is reasonable to expect, therefore, that the difficulty of connection to bundles of small diameter neuristors may not be a serious problem in non-digital computers. If high density logic capability requires non-digital techniques it is probable that neuristor bundles will be the better method of information transmission within high logic density computers.

This represents only a small area of potential neuristor applications. It is impossible to predict all the uses to which small diameter neuristors might be put. In modern technology, experience has shown that a successful device soon finds applications. Constantly more difficult requirements for heat dissipation, power consumption, miniaturization, environmental conditions and logical complexity make necessary the development of device components capable of multiple functions. The neuristor affords an intriguing realm of possibility as various combinations of self contained power source, delay line, amplifier, filter, discriminator circuit, and relay.

6.3.2. The Biological Effects of Strong Magnetic Fields

It has been tentatively suggested that high magnetic fields of strengths on the order of 25,000 to 100,000 gauss be associated with space craft for protection against high energy cosmic ray charged particles. Since space craft crews will be exposed to these high magnetic fields for long periods of time, it is reasonable to investigate the possible biological effects of such

exposure. A recent bibliography (Ref. 6-45) indicates that little relevant work has been done. Generally speaking, the fields are too weak and the duration is too short to be really significant. There is a basic reason for this shortage of data: maintenance of a 100,000 gauss field for a long period of time is expensive. This is especially so for whole body exposure of mammals.

The most relevant experiments seem to be those of Beischer (Ref. 6-46) whose findings are as follows:

"In a first laboratory study a variety of biological material has been exposed at the Naval Research Laboratory to magnetic fields of very high intensity and considerable gradient.

"Mice survived exposure to a practically homogeneous field of 120,000 gauss and to a field with a gradient of 7,000 gauss/cm. The experiment was terminated after one hour.

"Drosophila survived exposure to a homogeneous field of 115,000 gauss. A field gradient of more than 6,000 gauss/cm had lethal effects in exposure of Drosophila to fields from 40,000 to 100,000 gauss field strength.

"Preliminary observations point to effects of very strong magnetic fields on Photobacterium fischeri and on the development of sea urchin eggs."

Beischer attributes the lethal effects on Drosophila to the effect of nonuniform magnetic fields on the transport properties of paramagnetic and diamagnetic molecules. The absence of gross effects of the nonuniform fields on the reproduction system of Drosophila and the survival of mice under like conditions leads Beischer to conclude that the observed lethal effect is some very specific phenomenon unique to Drosophila.

The effect of strong magnetic fields on the development of sea urchin eggs indicates that magnetic fields may influence cell growth processes.

Beischer elsewhere (Ref. 6-47) has surveyed information on the effects of high magnetic fields on humans. The results are largely negative but the exposures were too brief or at too small field strengths to be convincing in the present context. Beischer also has suggested the possibility that absence of magnetic fields normally encountered by biological

organisms, i. e., geomagnetic fields, may also have biological effects.

In view of the general lack of data on magnetic field effects, it is desirable to develop a systematic program to study the problem.

Relevant biological damage may be divided into two categories which are (1) effects resulting in short term reduction of organism efficiency and (2) effects resulting in long term damage to the organism. These categories overlap: injury to and subsequent removal of an arm is an obvious example of damage in both categories.

Experiments to determine whether or not magnetic fields can cause serious biological effects fall into four categories.

1. Investigations of the effects of magnetic fields on the operating efficiency of organisms in both short run and long run exposures.
2. Investigations of organism changes resulting from exposures which, however, do not affect operating efficiency.
3. Physiological investigations of effects on particular tissues.
4. Basic research on the effect of the field on cytological processes such as cell membrane processes, neural conduction, etc.

The overall program would have two justifications. One is moral and economic. When an extremely expensive piece of equipment is to be piloted by human personnel on difficult missions it is both immoral and expensive to perform the ultimate experiment, namely, the efficiency effects of long term exposure, without prior experimentation of other types to justify the assumption of highly probable survival. The second justification is scientific in nature. Even if an astronaut were to perform a complicated mission of long duration in a strong magnetic field, whether on a space flight or in an earthbound simulation of such a flight, the results of the experiment could not really be interpreted without experiments in categories 2, 3 and 4. This would be true even if the long term experiment showed no efficiency reduction after long term exposure.

The reason for the importance of all four categories of experiments is that they differ in sensitivity. The categories are arranged in order of increasing sensitivity. This may be illustrated by nuclear radiation effects. One would expect to be able to observe changes in at least some individual tissues for lower dosages than are required to affect the blood

count. The blood count is, however, changed for radiation exposure much less than those required to reduce efficiency. Knowledge of detailed mechanisms, radiation damage; i. e., atomic displacements, would lead one to suspect possible long term cumulative effects, especially in genetics.

Category 1 experiments have generally produced negative results with the exception of experiments on Drosophila referred to above. Barnothy has reported positive results in Category 2 experiments, (quoted in Ref. 6-47) but Eiselern, et al., were unable to confirm these results, (quoted in Ref. 6-47). These experiments were all performed at relatively low field strengths and/or relatively short exposure times. The results are hardly conclusive.

Beischer's experiments on sea urchin eggs and on Bacterium fischeri produced positive results from magnetic field exposure. (Beischer actually urges some caution here because the temperature control of the experiment was not entirely satisfactory.) Although sea urchin eggs and Bacterium fischeri are, perhaps, more complex than "simple tissue", these experiments are very close to Category 3 experiments -- and positive results seem to be obtained even for short exposure.

Magnetic fields apparently can have effects at the cellular level. The inevitable question is whether or not these effects could turn out to be significant to men exposed for long periods of time to high magnetic fields. To define what is going on it is obviously desirable to go further and perform experiments on as simple tissue configurations as possible. This consideration brings the discussion back to nerve axons for the same reasons discussed earlier. The axon is a geometrically simple cell component embodying many of the fundamental activities of the cell. The value of axon experiments in understanding, for example, membrane behavior has already been discussed at length. There is additional reason to be interested in the axon for magnetic field study. Although vector concepts seem necessary to describe axon impulse conduction one would not necessarily expect this a priori. The magnetic field is, however, inherently and fundamentally a vector phenomenon. Experiments should reflect this and perhaps the best biological sample with which to do this is the nerve axon. Any effect on the neural process would be critical to overall organism function. The remainder of this section will discuss some of this work, useful experiments which should be performed and some of the physical theory involved.

There has been little work on the effects of magnetic fields

on nerve axon behavior. Liberman, Vaintsvarg and Trofina (Ref. 6-48) studied the effect of a constant magnetic field on the excitation threshold of isolated frog nerve. The magnetic field was 10,000 gauss perpendicular to the nerve axis. These investigators found no effects for thresholds measured with an accuracy of $\pm 1\%$. This negative result contradicted positive findings of earlier Russian investigators (quoted in Ref. 6-48).

The field strength of the above experiment is too low for the results to be considered as decisive. There is another, more fundamental criticism. The transverse field may be of interest for some types of physical phenomena but it is clear that the radial directions are likely to be of more interest. The radial direction has a fixed relationship only to longitudinal magnetic fields. It seems reasonable to assert, therefore, that negative results should not be accepted as final until the effect of longitudinal fields has been investigated.

A longitudinal field can be established with a solenoid. It should not be difficult to mount nerve specimens with necessary excitation and detection electrodes along the axis of a solenoid which yields high field strengths. Careful cooling would be required.

Once an operating apparatus is set up obvious investigations would include the effect of longitudinal magnetic fields on excitation thresholds, action potential conduction velocity and chronaxie. Since vector quantities are involved, questions of electrode configuration as discussed in Section 6.2.1 above should also be considered.

The question naturally arises, what physical phenomena might result in magnetic field effects on nerve axon behavior. Beischer refers to and rejects effects on magnetic moments of electrons (Ref. 6-46). Liberman, et al., reject Hall effects on ionic flows on the grounds that the ionic mobilities are too low, but suggest the possibility that electronic conduction may occur within the membrane (Ref. 6-48). This is not inconsistent with the electron transfer mechanisms discussed by Szent Gyorgi (Ref. 6-42). Heinmetz and Hirschman believe that magnetic fields will not significantly affect flow of ions but that combinations of changing electric and magnetic fields may significantly affect ionic flow and the behavior of dipoles (Ref. 6-49). The general opinion appears to be that there are phenomena which could be significant but the significance appears to be marginal at best.

In Section 6.2.2 of this report, a statistical theory of the membrane was introduced. The significance of this theory for magnetic field effects will now be discussed.

The essential characteristic of the statistical membrane theory is that the sodium current through the membrane is an essentially improbable process. Spurts of ions pass through the membrane at widely separated points. Nachmanson's work indicates that each spurt consists of about 10^3 sodium ions.

Ringer's solution contains the equivalent of approximately 10 grams of NaCl per 1000 cm^3 . This corresponds to a sodium ion concentration of 1.03×10^{20} ions per cm^3 . The 1000 ions in one spurt occupy a volume of $9.7 \times 10^{-18} \text{ cm}^3$ which would be a hemisphere of approximate radius $1.67 \times 10^{-6} \text{ cm}$. This is illustrated in Fig. 6-30. The membrane is shown with a channel through it. This is the channel opened up by the release of acetylcholine. Its diameter is the order of 20\AA . It is assumed that the ions of the spurt are from a hemisphere volume as shown. Approximately half of the ions are located outside of a hemisphere of 130\AA . During the spurt the ions must travel approximately $3.46 \times 10^{-6} \text{ cm}$. Since the spurt duration is about $0.25 \times 10^{-3} \text{ sec}$, the average ion velocity is $1.38 \times 10^{-2} \text{ cm/sec}$. But the channel represents a restriction of the flow channel by a factor of about 50. The velocity through the channel should be about 0.7 cm/sec .

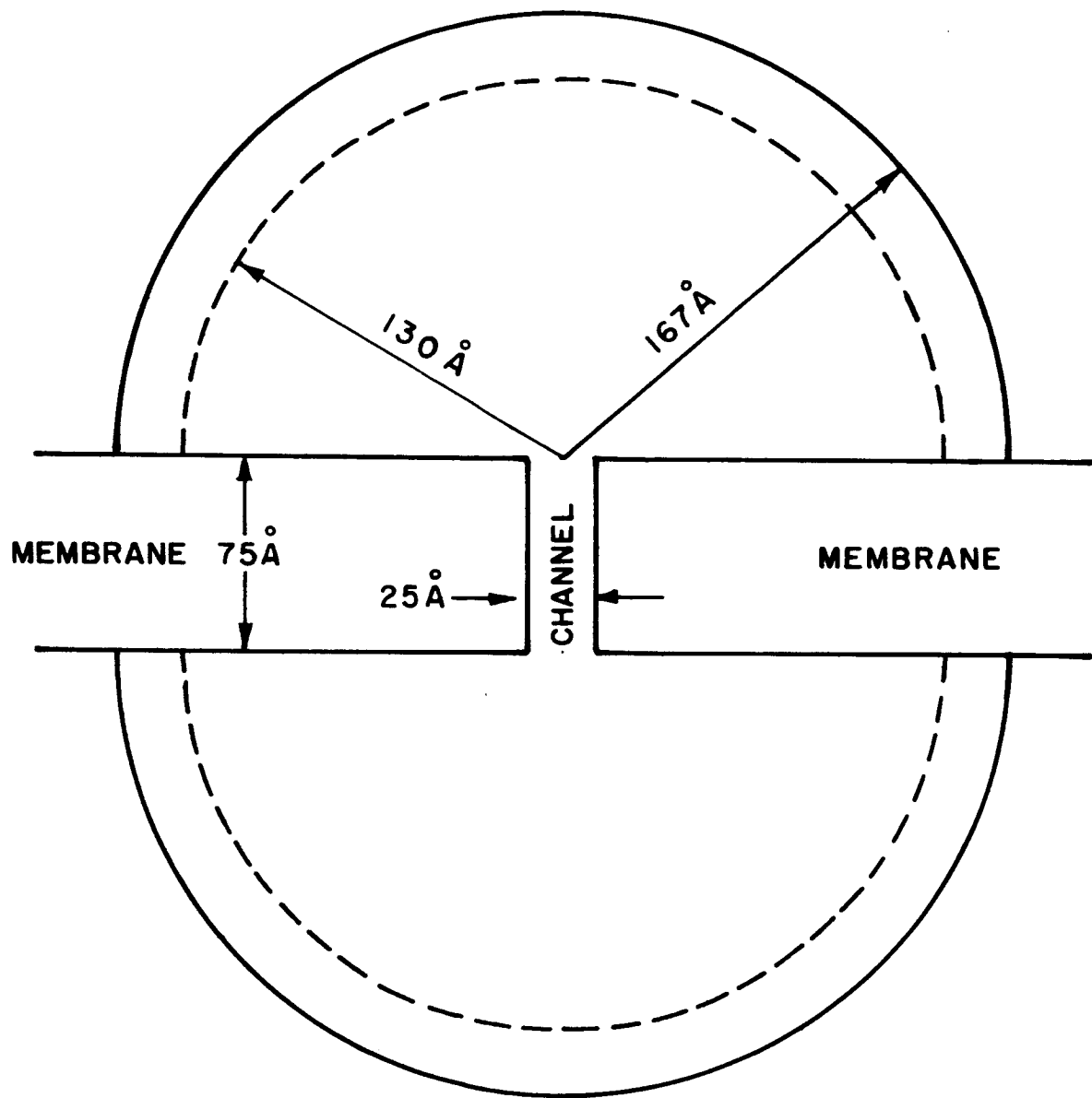
If the velocity in the channel is actually 0.7 cm/sec , then 10^3 ions will pass through the channel in $0.25 \times 10^{-3} \text{ sec}$, if the distance between ions in the membrane is $0.175 \times 10^{-6} \text{ cm}$. That is, there are about four ions simultaneously in the channel whose volume is about $3 \times 10^{-20} \text{ cm}^3$. The volume for each ion is about $0.75 \times 10^{-20} \text{ cm}^3$. This is almost equal to the volume per ion in the original solution. Debye has calculated (Ref. 6-52) that the electric field $0.1 \times 10^{-6} \text{ cm}$ from a singly charged ion in a solution is 100,000 volts/cm. It is not likely that the distance between ions in the channel will be much less than the $0.175 \times 10^{-6} \text{ cm}$ calculated above.

It appears, therefore, that the ion velocity in the membrane channel is the order of 1 cm/sec , if the statistical theory of the membrane is correct. The question arises, how will ions of such velocities interact with a magnetic field parallel to the nerve axis?

The radius of curvature of a charged particle in a magnetic field perpendicular to the velocity is given in rationalized MKS units by:

FIG. 6-30

MEMBRANE SODIUM ION CHANNEL



$$R = \frac{mv}{Bq} \quad (6-38)$$

R is radius in meters

m is mass in kilograms

v is the magnitude of the velocity in meters per second

B is magnetic field in webers /m²

q is charged in coulombs.

For a sodium ion, $m = 3.84 \times 10^{-26}$ kg. and $q = 1.6 \times 10^{-19}$ coulomb and Equation 6.38 becomes:

$$R = 2.4 \times 10^{-7} \frac{v}{B} \quad (6-39)$$

If $v = 10^{-2}$ m/sec or 1 cm/sec

and $B = 10$ webers/m² or 10^5 gauss,

then $R = 2.4 \times 10^{-10}$ meters, or 2.4 \AA .

That is, R can easily be the same order of magnitude as the dimensions of the channel in the membrane. Actually the ions will tend to be attached at least to H₂O molecules. This will tend to increase m, and therefore, R, thus strengthening the argument.

These calculations are all very crude but it seems reasonable that for low magnetic fields the ionic radius of curvature will be appreciably greater than channel dimensions and for very large magnetic fields the ionic radius of curvature would, in principle, be very much less than channel dimensions. Approximate equality will occur for magnetic fields in the range of 10^4 to 10^5 gauss.

It seems reasonable to assume that in either case of inequality the magnetic field effects will simply be those associated with ordinary ionic flow. In the cases for which the radius of curvature is the same order of magnitude as the channel dimensions, interesting interactions may be expected to occur. Such graphic pictures as ions "slamming into the channel walls" will correspond to aspects of physical reality.

Before dropping the question of magnetic fields it is appropriate to investigate magnetic field effects in ordinary ionic conduction. The force on a charge moving under the influence of an electric field in the presence of a magnetic field is given in rationalized MKS units by

$$\vec{F} = q (\vec{E} + \vec{v} \times \vec{B}) \quad (6-40)$$

where the vector notation is to be noted.

The velocity of an ion in an electric field is given by

$$\vec{v} = \mu \vec{E} \quad (6-41)$$

where μ is the mobility. In rationalized MKS units, $\mu \approx 10^{-7}$ for typical ionic solutions. Equation (6.40) then becomes

$$\vec{F} = q (\vec{E} + 10^{-7} \vec{v} \times \vec{B}) \quad (6-42)$$

This equation strictly speaking applies only if \vec{v} is parallel to \vec{E} . It is clear that \vec{B} must be the order of 10^7 webers/m² for the magnetic term of this equation to be significant. Such magnetic fields are not likely to be encountered and have not been produced. This is undoubtedly the reason that the Hall effect has not been observed in ionic solutions.

This discussion may be summarized by the following statements:

1. Magnetic field effects have been observed on the cellular level.
2. Standard physical considerations appear to give no, or generally unsatisfactory, explanations of this.
3. The statistical theory of the membrane discussed in Section 6.2.2. does yield a possible explanation.
4. Nerve axon experiments using longitudinal magnetic fields of strengths the order of 10^4 to 10^5 gauss should yield interesting results - either positive or negative.

6.4. Summary

This discussion of neural phenomenon has been written with the twofold objective of giving an organized account of biological knowledge and of indicating possible applications to the engineering of sensor associated transmission and information handling of biological concepts.

The biological discussion begins with a survey of the physiology of nerve axons. The importance of the cylindrical geometry of axons for experimental investigation of cell processes is emphasized. Physiological experiments on the nerve axon are reviewed and the fundamental role of the membrane in axon impulse conduction is demonstrated.

Experiments on the excitation of nerve axons are reviewed. Results of these experiments are shown to be generally consistent with the logically simpler experiments discussed earlier. The significance of longitudinal electric fields in axon excitation is discussed. Recent preliminary experiments at Allied Research Associates suggest that the longitudinal electric fields are important in axon behavior.

The membrane is discussed. The fundamental work of Hodgkin and Huxley on the relationship between depolarization and membrane permeability is compared to Nachmanson's work on the effect of the acetylcholine cycle on membrane permeability. A fundamental statistical theory of the membrane is introduced which establishes the connection between these two viewpoints and suggests further confirmatory experiments.

Attention is then shifted to the general organization of the nervous system. A survey of certain psychological and philosophical studies suggests introduction of "a fundamental theorem of nervous system organization". This theorem, which awaits proof, states that an optimum design of a pattern recognition device is necessarily composed of both digital logical and perceptron components. A practical engineering application, the recognition of printed letters and numerals is suggested as an excellent opportunity for investigation of this theorem.

Attention is then directed to the comparison between neuristors and coaxial cables. It is concluded that for materials of comparable conductivity, the information transmission rate of coaxial cables is comparable to that of neuristor bundles of equal total diameter for distances of approximately three feet. Possible applications of neuristors are discussed against the background of this result and in terms of their potentially unique characteristics.

Biological effects of strong magnetic fields are surveyed. It is shown that existing data are inadequate to predict effects of high magnetic fields for long term exposures. A brief outline is given of certain experiments on nerve axons which would be especially significant. The statistical theory of the membrane, introduced earlier, is used to derive a possible magnetic field effect on cell processes.

In a subject of such vast proportions it is impossible to present a complete picture. It is believed that by careful organization of previous results, by extensions of previous theory and by attention to engineering applications growing logically from biological fundamentals it will be possible to achieve an integration of considerable bionic significance.

REFERENCES - SECTION 6

- 6-1 Robertson, J. D. "New Unit Organelle of Schwann Cells", in Biophysics of Physiological and Pharmacological Actions. Publication No. 69 of American Association for the Advancement of Science, Washington, D.C., (1961)
- 6-2 Gray, Henry, Anatomy of the Human Body. Gross, C. M., ed., Lea and Febiger, Philadelphia, (1948)
- 6-3 Robertson, J. D. "The Molecular Biology of Cell Membranes", in Molecular Biology. Nachmansohn, D., ed., Academic Press, New York, (1960)
- 6-4 Gengerelli, J. A., Hotter, N. J., and Glasscock, W. R., "Magnetic Fields Accompanying Transmission of Nerve Impulses in the Frog's Sciatic Nerve", J. Psychol. 52, 317, (1962)
- 6-5 Brazier, Mary A. B. "The Electrical Activity of the Nervous System", MacMillan, New York, (1960)
- 6-6 Baker, P. F., Hodgkin, A. L., and Shaw, T. I. "Replacement of the Axoplasm of Giant Nerve Fibres with Artificial Solutions", J. Physiol. 164, 330, (1962)
- 6-7 Tasaki, I. "Initiation and Abolition of the Action Potential of a Single Node of Ranvier", J. Gen. Physiol. 39, 377, (1956)
- 6-8 Cole, K. S. "Ions, Potentials and the Nerve Impulse", in Electrochemistry in Biology and Medicine. Shedlovsky, T., ed., John Wiley, New York, (1955)
- 6-9 Abbott, B. C. "Energetics of Nerve Activity", in Biophysics of Physiology and Pharmacological Actions. American Association for the Advancement of Science, Publication No. 69, Washington, D.C., (1961)
- 6-10 Hodgkin, A. R. and Huxley, A. F. "A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve", J. Physiol. 117, 500, (1952)
- 6-11 Tasaki, I. "Conduction of the Nerve Impulse", in Handbook of Physiology. Section I, Vol. I, Field, J., ed., American Physiological Society, Washington, D.C., (1959)

- 6-12 Katz, B. "Action Potentials from a Sensory Nerve Ending", *Journal Physiology (Lond.)*, 111, 248, (1950).
- 6-13 Koketsu, K. "Mechanism of Active Depolarization-Dispensability of Sodium", in Biophysics of Physiological and Pharmacological Actions.
- 6-14 Huxley, A. F. "Ion Movements During Nerve Activity", *Annals of New York Academy of Sciences*, 81 (2), 221, (1959)
- 6-15 Rushton, W. A. G. "The Effect Upon the Threshold for Nervous Excitation of the Length of Nerve Exposed and the Angle Between Current and Nerve", *J. Physiol.* 63, 357, (1927)
- 6-16 Rushton, W. A. H. "Identification of Lucas' α Excitability", *J. Physiol.* 75, 445, (1932)
- 6-17 Tuttle, W. W. and Schotteluis, B. A. Physiology Laboratory Manual. C. V. Mosby Company, St. Louis, (1963) p. 123.
- 6-18 von Neumann, J. The Computer and the Brain. Yale University Press, New Haven, (1963)
- 6-19 Schmitt, O. H. Electrochemistry in Biology and Medicine. Shedlovsky, Theodore, ed., John Wiley, New York, (1955)
- 6-20 Healer, J., Arms, A., Merchant, J., and Britt, E. Summary Report on a Review of Biological Mechanisms for Application to Instrument Design. ARA 1025, Allied Research Associates, Inc., January, (1962)
- 6-21 Nachmansohn, D. Chemical and Molecular Basis of Nerve Activity. Academic Press, New York, (1959)
- 6-22 Tasaki, I. and Spyropoulos, Constantine S. Influence of Temperature on Biological Systems. Johnson, F. H., ed., American Physiological Society, Washington, D.C., (1957) p. 201.
- 6-23 Granit, R. Receptors and Sensory Perception. Yale University Press, (1963)
- 6-24 Chauchard, P. The Brain. Grove Press Inc., New York, (1962)
- 6-25 von Neumann, J. The Computer and the Brain. Yale University Press, New Haven, (1963)
- 6-26 MacKay, D. M. Mechanization of Thought Processes. Her Majesty's Stationery Office, London, (1962)

- 6-27 The Perceptron: A Theory of Statistical Separability in Cognitive Systems. ASTIA Document No. 155059, Cornell Aeronautical Laboratory, Buffalo, New York, (1958)
- 6-28 Hadamard, J. The Psychology of Invention in the Mathematical Field. Dover Publications, New York, (1954)
- 6-29 Kubie, L. Neurotic Distortion of the Creative Process. Noonday Press (a subsidiary of Farrar, Straus and Cudahy), (1961)
- 6-30 Wiener, N. Cybernetics. John Wiley, New York, (1961)
- 6-31 Poincare, H. The World of Mathematics. Vol. IV, Newman, James, R., ed., Simon and Schuster, New York, (1956)
- 6-32 Rosenfeld, L. Physics Today , 16, 47, October, (1963)
- 6-33 Abegg, L. The Mind of East Asia. Thames and Hudson, London - New York, (1952)
- 6-34 Schiff, L. I. Quantum Machanics. McGraw-Hill, New York, (1955)
- 6-35 Heisenberg, U. Physics and Philosophy. Harpers and Brothers, New York, (1958)
- 6-36 Bohr, N. Atomic Physics and Human Knowledge. John Wiley, New York, (1958)
- 6-37 Pahl, P. M. and Johnson, D. I. Linguistic and Engineering Studies in the Automatic Translation of Scientific Russian into English. (Technical Report Phase II) Contract AF 30(602)-1827, U. of Wash. Press, Seattle, (1960)
- 6-38 Block, H. D. "The Perceptron: A Model for Brain Functioning, I and II", Reviews of Modern Physics, 34, 123, January, (1962)
- 6-39 Rosenblatt, F. "Two Theorems of Statistical Separability in the Perceptron", in Mechanization of Thought Processes. NPL Symposium No. 10, Her Majesty's Stationery Office, London, (1962)
- 6-40 Ruth, S. B. "The Hybrid Computer - End of an Argument", Electronic Industries, 22, 38 Dec. (1963).
- 6-41 Hubel, D. H. , "The Visual Cortex of the Brain", Scientific American, 209, 54 Nov. (1963).
- 6-42 SzentGyorgyi, A., Introduction to a Sub Molecular Biology, Academic Press, N. Y. 1960).

- 6-43 Crane, H.D. , "Neuristor - A Novel Device and System Concept,"
Proc. IRE, 50, 2048, October (1962)
- 6-44 Reference Data for Engineers, 4th Edition, International Telephone
and Telegraph Corp. , New York, N. Y. , (1957).
- 6-45 Davis, L.D. , Pappajohn, K. , and Plavniecks, I. M. , "Bibliography
of the Biological Effects of Magnetic Fields," Proc. Fed. Amer.
Soc. Exptl. Biol. , 25 (5 Pt 2 Suppl. 12): 1-38 (1962)
- 6-46 Beischer, D. E. , Survival of Animals in Magnetic Fields of 120,000
Gauss, Research Report of U.S. Naval School of Aviation Medicine,
Pensacola, Florida, 25 July (1962).
- 6-47 Beischer, D. E. , "Human Tolerance to Magnetic Fields," *Astronautics*
7, 24, March (1962).
- 6-48 Liberman, E.A. , Vaintsvarg, M.N. , and Tsofina, L.M. , "The
Effect of a Constant Magnetic Field on the Excitation Threshold of
Isolated Frog Nerve," *Biofizika* 4, 505-506, (1959).
- 6-49 Heinmetz, F. , and Hirschman, A. , "Considerations on the Effects
Produced by Superimposed Electric and Magnetic Fields in Biological
Systems and Electrolytes," *Phys. Med. and Biol.* 5 (3), 271-287
(1961).
- 6-50 Robertson, J.D. , *J. Biophys. Biochem. Cytol.* 3, 1043-1048 (1957).
- 6-51 Danielli, J.F. , and Davson, H. , *Journal Cellular Comp. Physiol* 5, 495
(1935).
- 6-52 Debye, P. , Polar Molecules, Dover Publications, N. Y. (1929)
- 6-53 Fischer, G. L. , Jr. , Ed. , Optical Character Recognition, Spartan
Books, Washington, D. C. , (1962).